Welcome To

ISSCC 2014
Session 10
Mobile Systems-on-Chip (SoCs)

A 28nm DSP Powered by an On-Chip LDO for High-Performance and Energy-Efficient Mobile Applications

Martin Saint-Laurent, Paul Bassett, Ken Lin, Yuhe Wang, Son Le, Xufeng Chen, Maen Alradaideh, Tom Wernimont, Kartik Ayyar, Dan Bui, Dwight Galbi, Allan Lester, Willie Anderson

QUALCOMM®

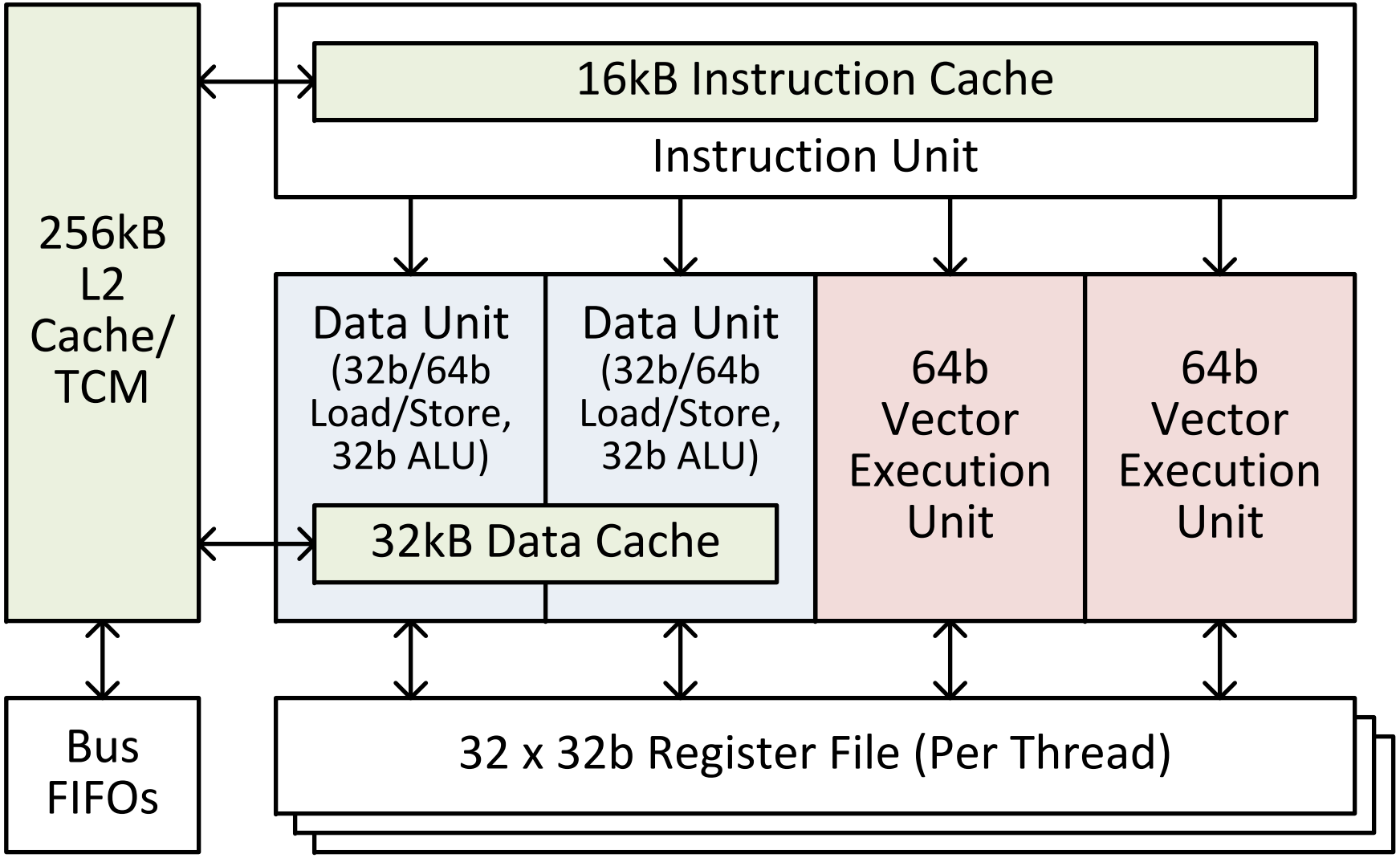
Outline

- Introduction
- Core Power Gating and Voltage Scaling
- Low-Power Clocking Techniques
- Asynchronous FIFOs and Arrays
- Core Frequency/Power Data
- Conclusions

Hexagon DSP Overview

- Targets high performance and low power across a wide variety of multimedia and modem applications, under aggressive area targets.
- Pursues high IPC as opposed to high frequency: 4.7 DMIPS/MHz.
- Innovates in ISA.
- Multi-threaded to reduce L2 cache miss penalty without the need to make it large.
- Supports advanced power management.

Architecture



Technology Overview

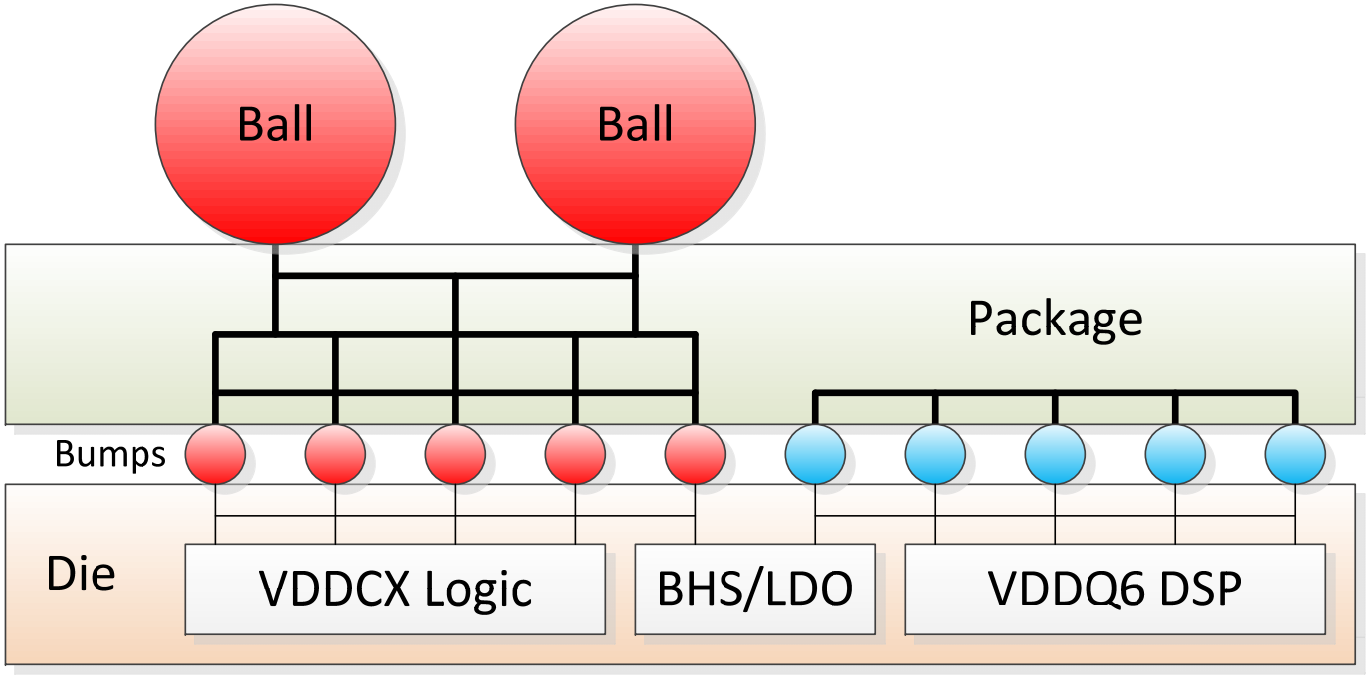
- 28nm High-k Metal-Gate Bulk Technology
- 3 Logic Threshold Voltages
- 3 Channel Lengths for Each Vt
- 8 Metal Layers:
  - M1-M6: 1X (90nm)
  - M7: 4X (360nm)
  - M8: 8X (720nm)
- 6T SRAM Cell: 0.127µm²

Outline

- Introduction
- Core Power Gating and Voltage Scaling
- Low-Power Clocking Techniques
- Asynchronous FIFOs and Arrays
- Core Frequency/Power Data
- Conclusions

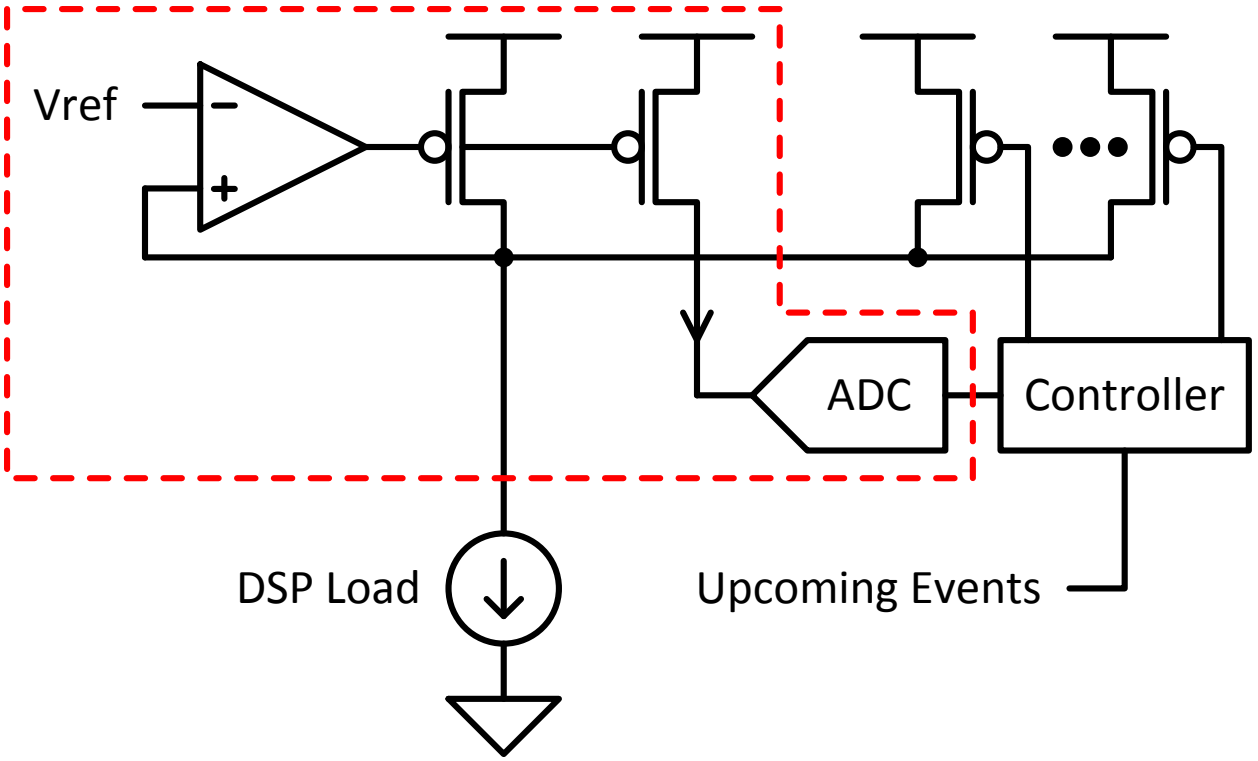
Core Power Gating

- A block head switch is connected in parallel with an LDO to produce the DSP voltage (VDDQ6).
- VDDQ6 is routed over the DSP using the package M1 and then connected to an on-chip power grid.



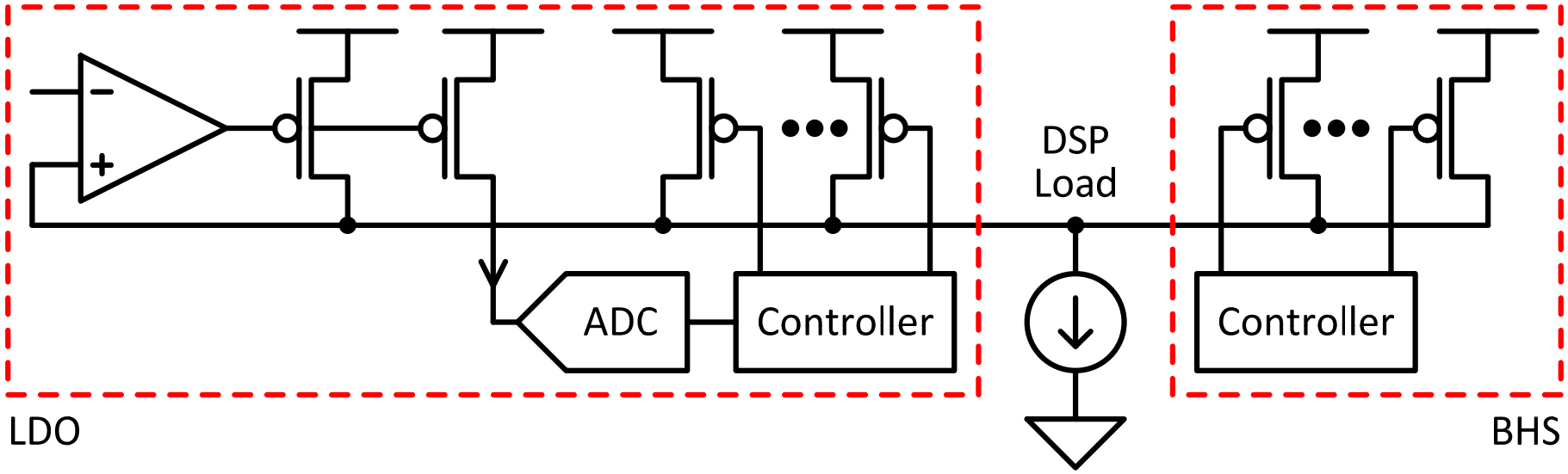
On-Chip LDO

- No external capacitor.
- The analog loop has a current mirror driving an ADC.
- The excess leakage is shifted to the digital loop.



LDO/BHS Transitions

- Allowed while the DSP is running.
- LDO to BHS: Progressively turn on the BHS to pinch the headroom of the LDO to nearly zero.
- BHS to LDO: Digitally force the LDO to its minimum impedance. Turn off the BHS. Enable regulation.

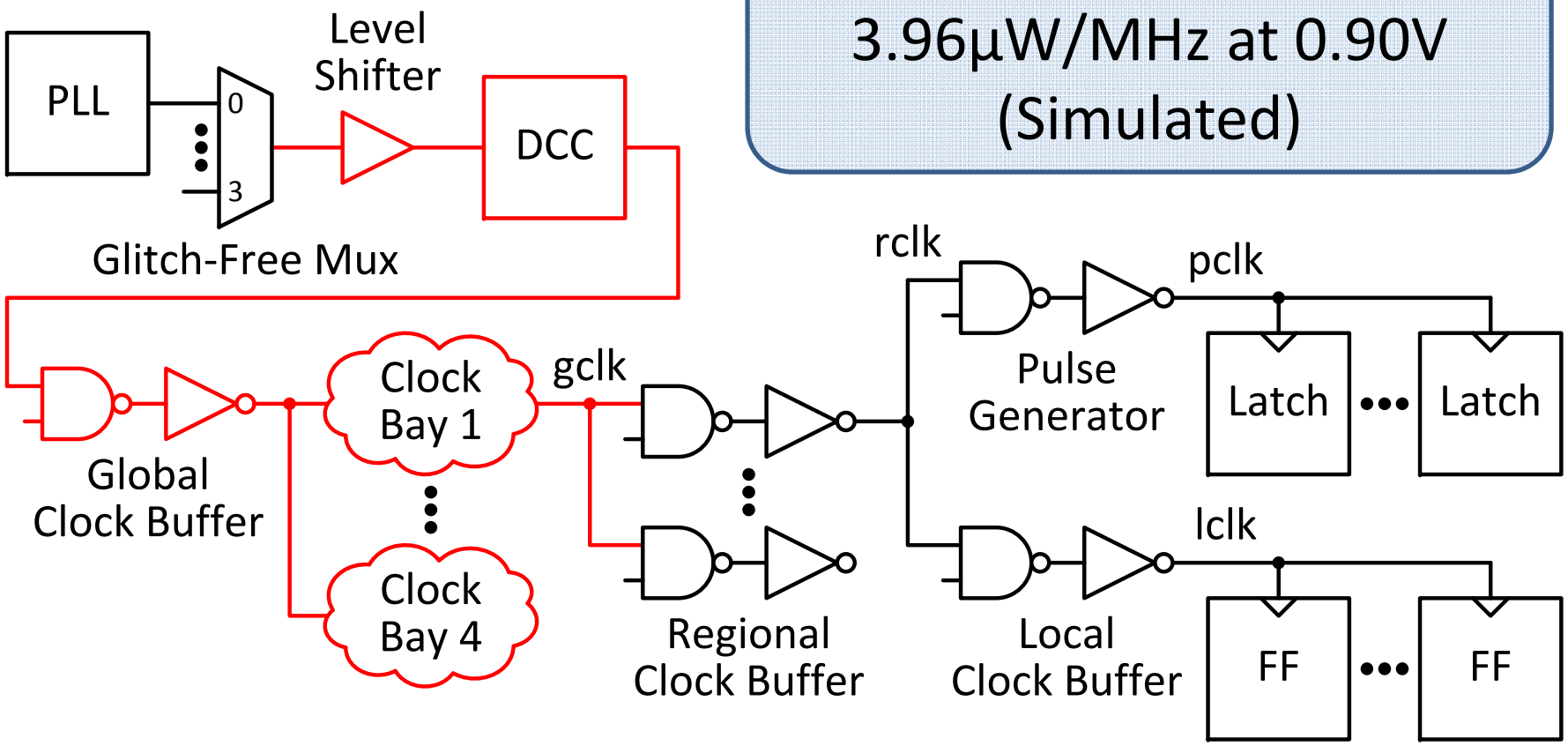


Outline

- Introduction
- Core Power Gating and Voltage Scaling
- Low-Power Clocking Techniques
- Asynchronous FIFOs and Arrays
- Core Frequency/Power Data
- Conclusions

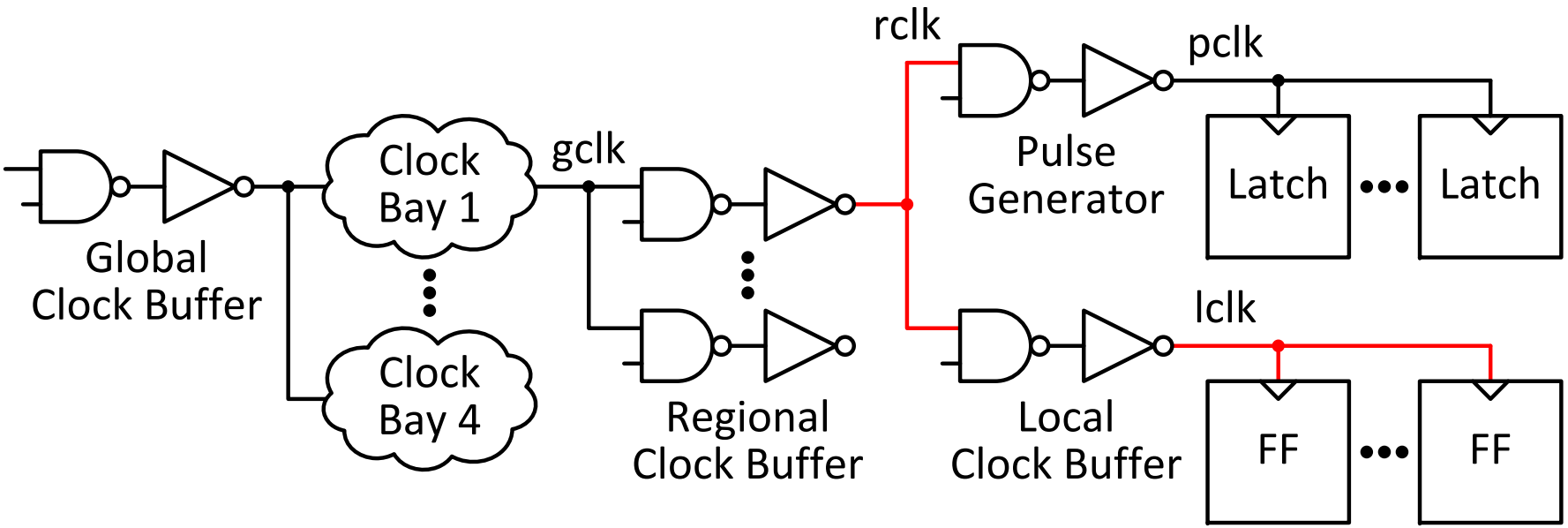
Clock Distribution

Dynamic Power:
3.96µW/MHz at 0.90V
(Simulated)

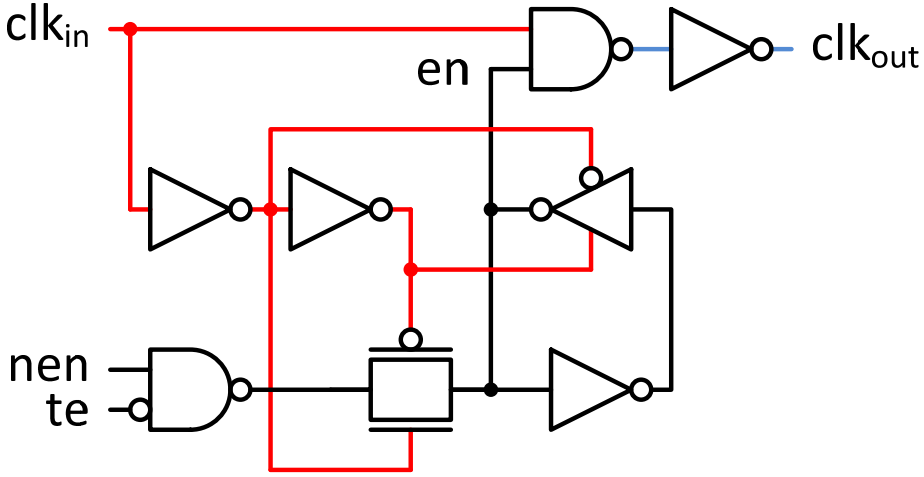


Clock Distribution

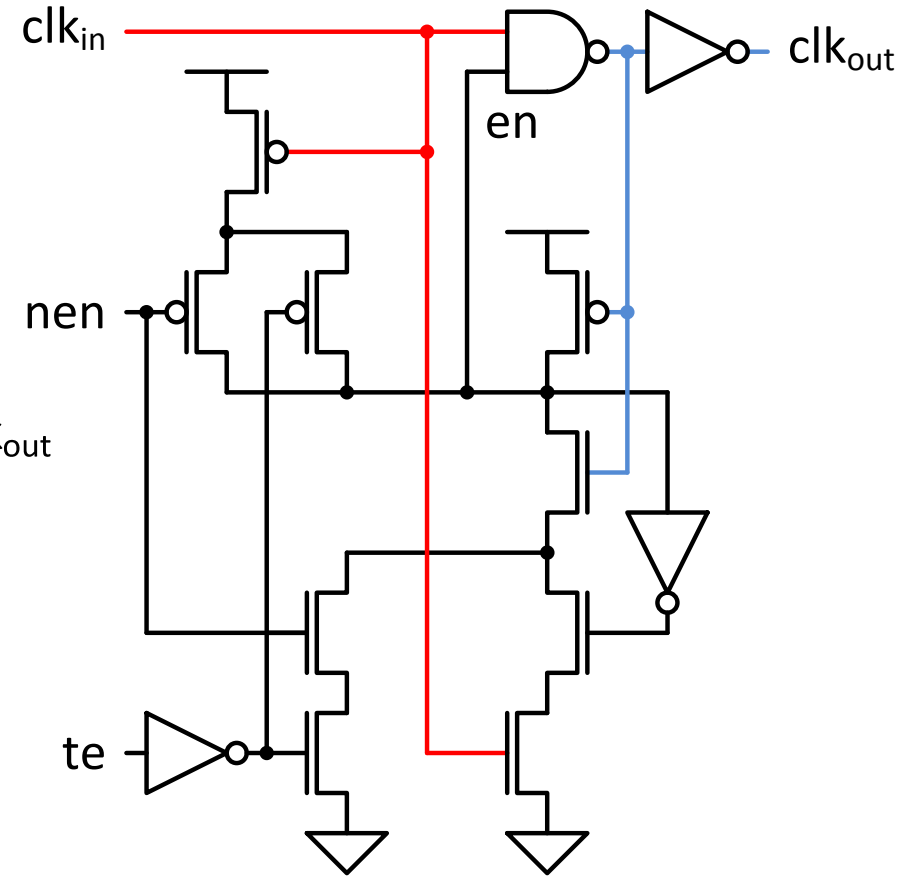
Typically, 80% of the regional clocks and 94% of the local clocks are gated off.



Low-Power Clock Gating Cell (CGC)



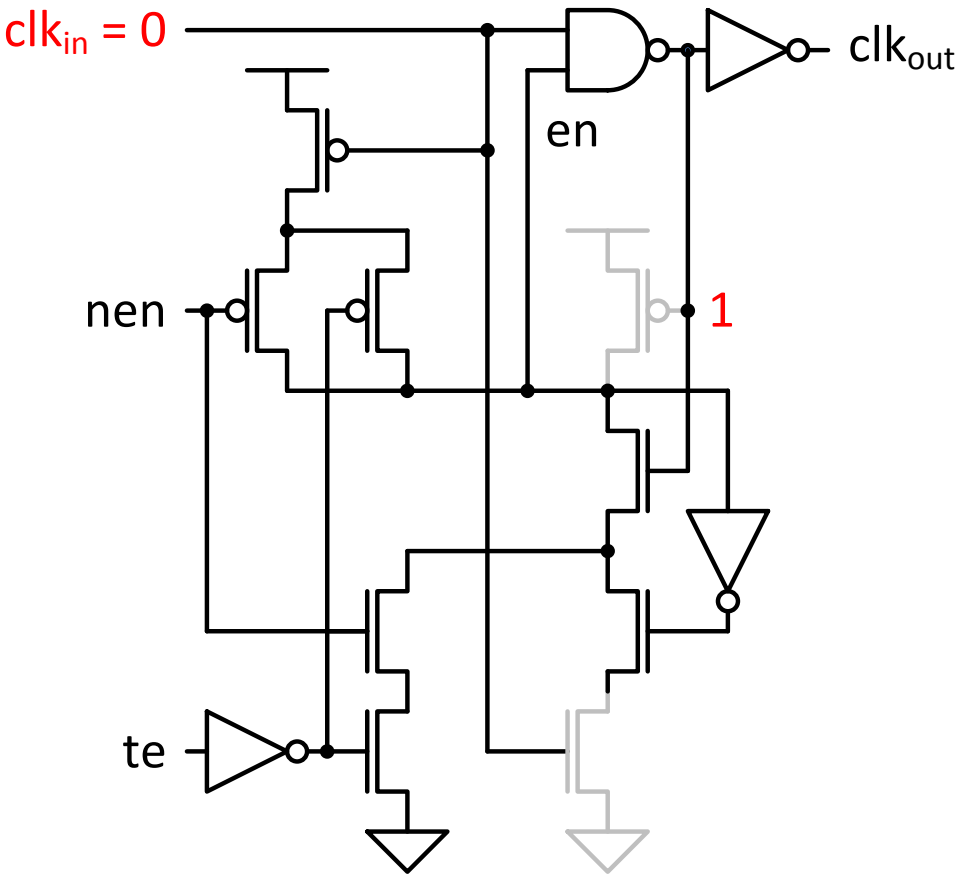
Conventional: 24T



Low Power: 19T

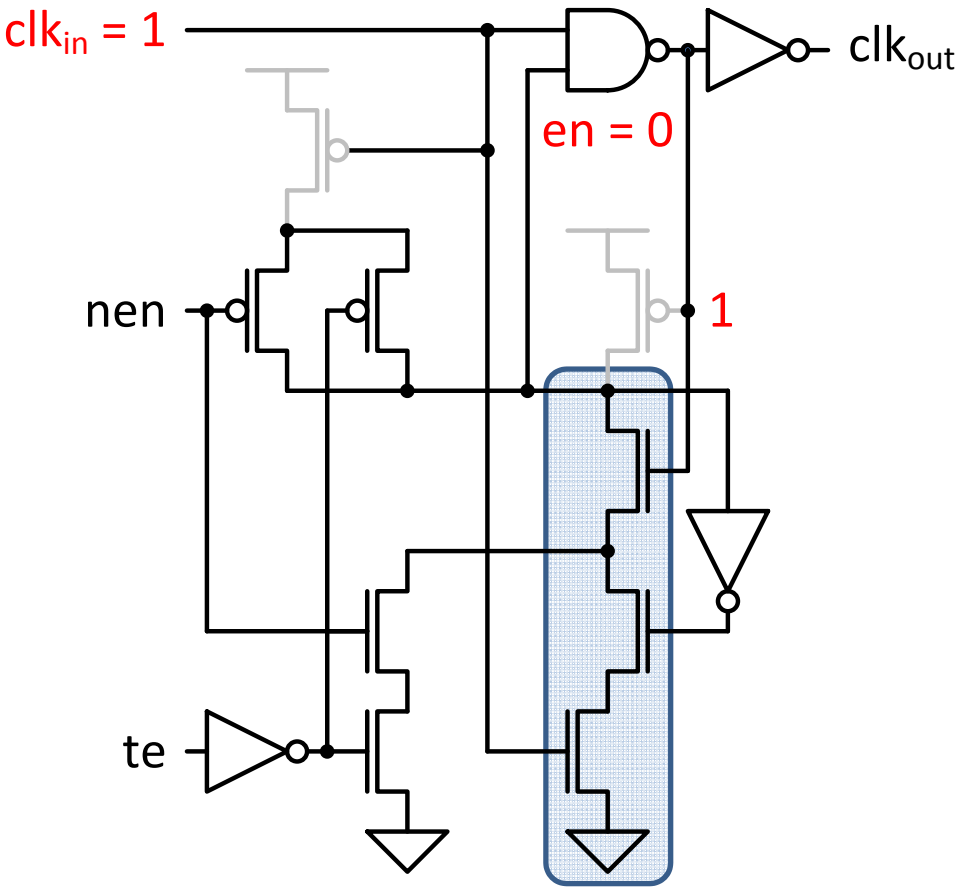
CGC Operation

- When clk_{in} = 0, the latch is transparent.
- When clk_{in} = 1, the latch is opaque.
- The NAND gate implements half of the keeper.



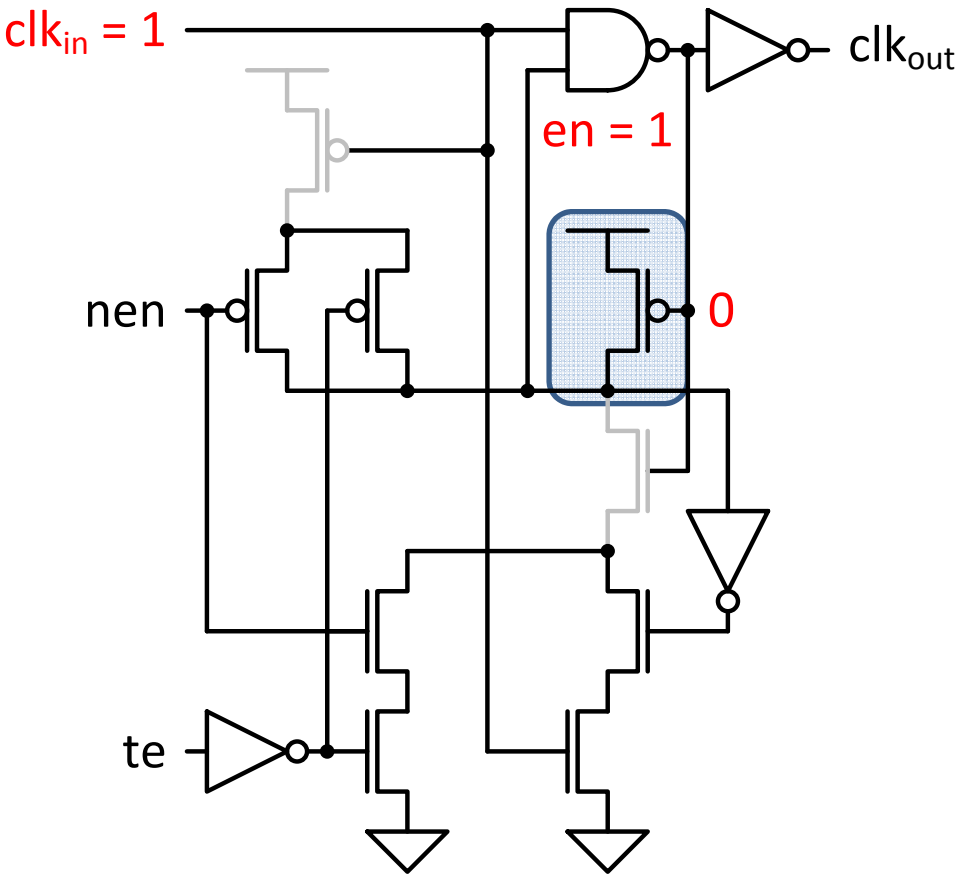
CGC Operation

- When clk_{in} = 0, the latch is transparent.
- When clk_{in} = 1, the latch is opaque.
- The NAND gate implements half of the keeper.



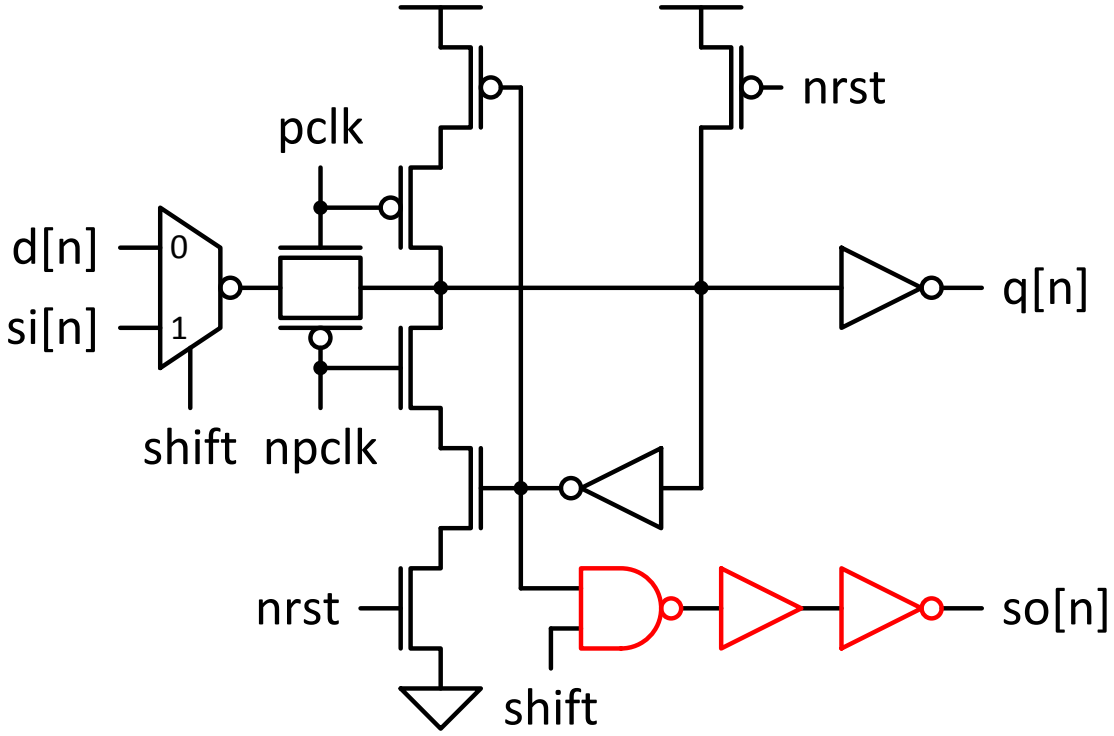
CGC Operation

- When clk_{in} = 0, the latch is transparent.
- When clk_{in} = 1, the latch is opaque.
- The NAND gate implements half of the keeper.



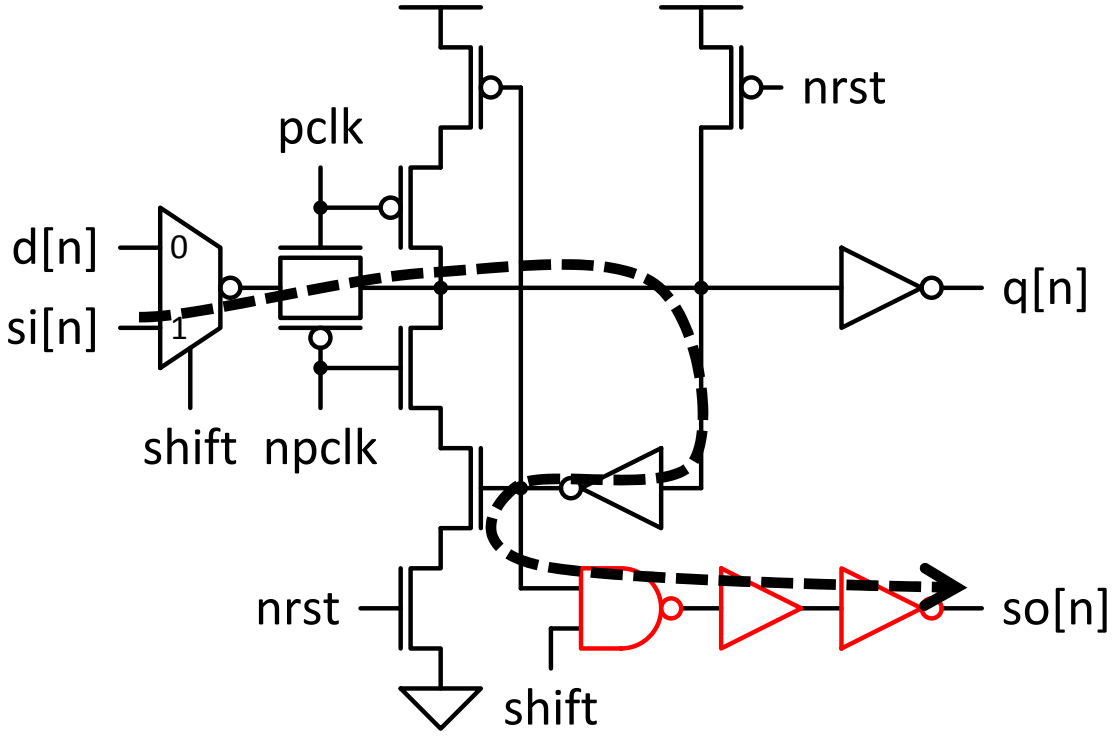
Pulsed Latches

- Up to 32 bits are combined in a single cell.
- Reduces variability due to layout context effects.
- Pulse width: 5 inversions.



Pulsed Latches

- Path from scan in to scan out: 6 inversions.
- The scan chain uses long-channel high-Vt gates.
- Must optimize using statistical techniques.

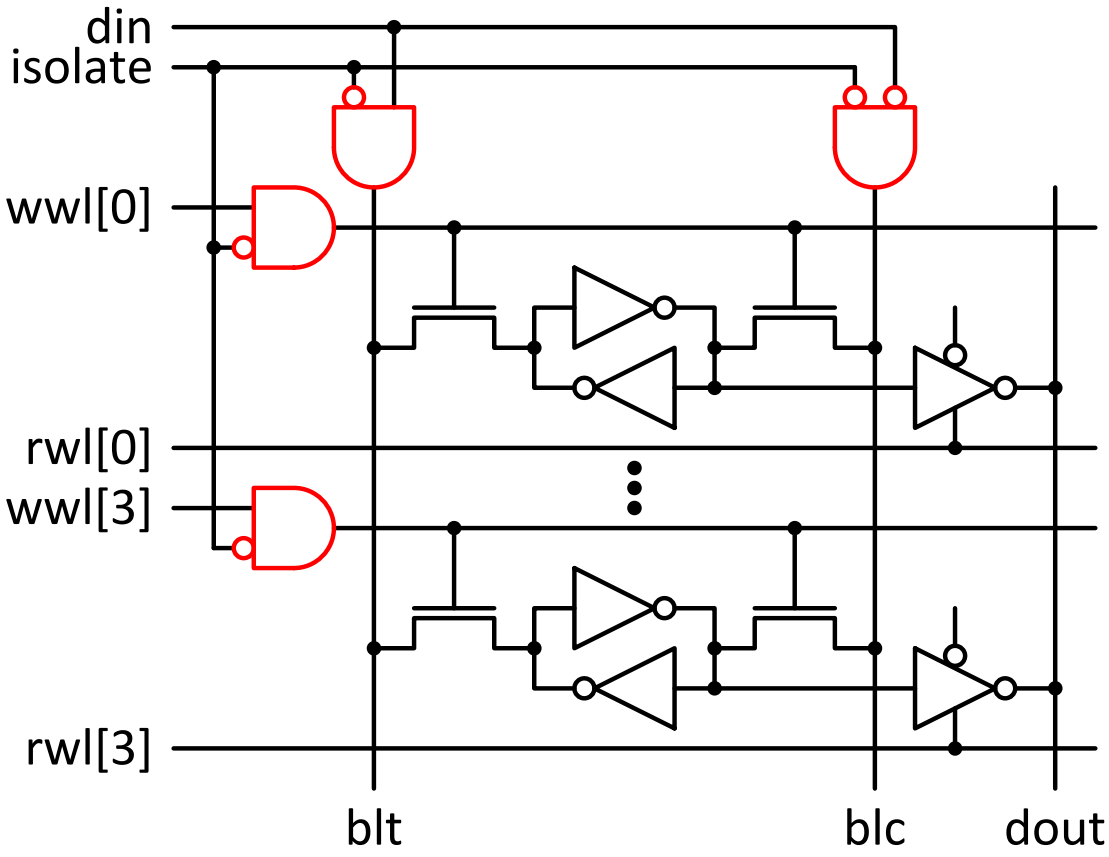


Outline

- Introduction
- Core Power Gating and Voltage Scaling
- Low-Power Clocking Techniques
- Asynchronous FIFOs and Arrays
- Core Frequency/Power Data
- Conclusions

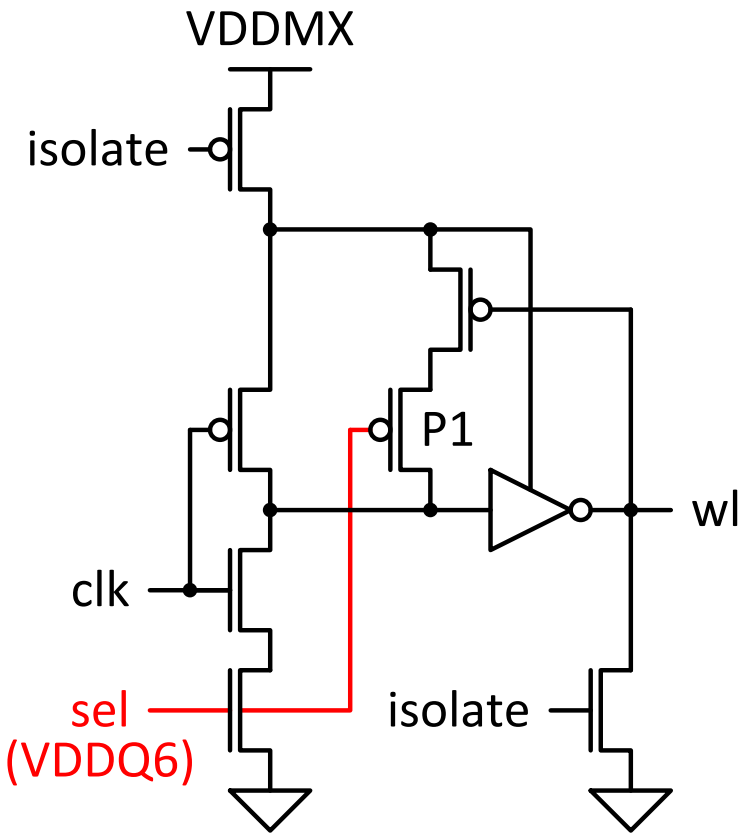
Asynchronous FIFOs

- Cross voltage and asynchronous clock boundaries.
- The write clock and data paths go through matched level shifter and isolation cells.
- The read/write pointers are controlled using synthesized logic.



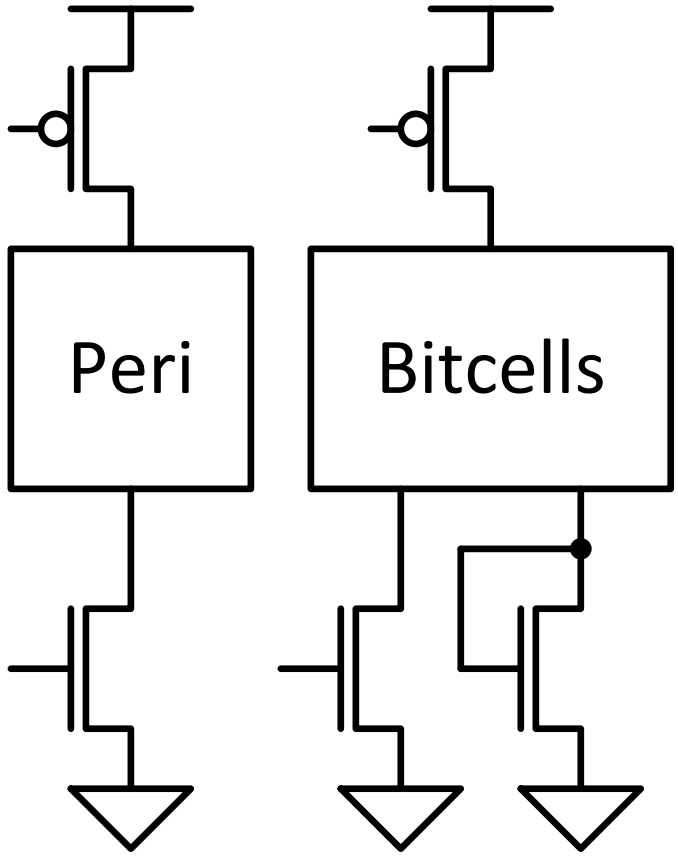
6T Arrays

- The 6T bitcells are on a dedicated memory supply (VDDMX).
- The address decoder is on the core supply (VDDQ6).
- Dynamic level shifters are embedded with the wordline drivers.
- P1 can be weakly on when sel = 1.
- The wordline drivers can be power gated.



6T Arrays

- Independent power switches for the bitcells and the periphery.
- Source biasing to reduce the retention leakage.
- Track the delay of the level shifters when firing the sense amps.



Outline

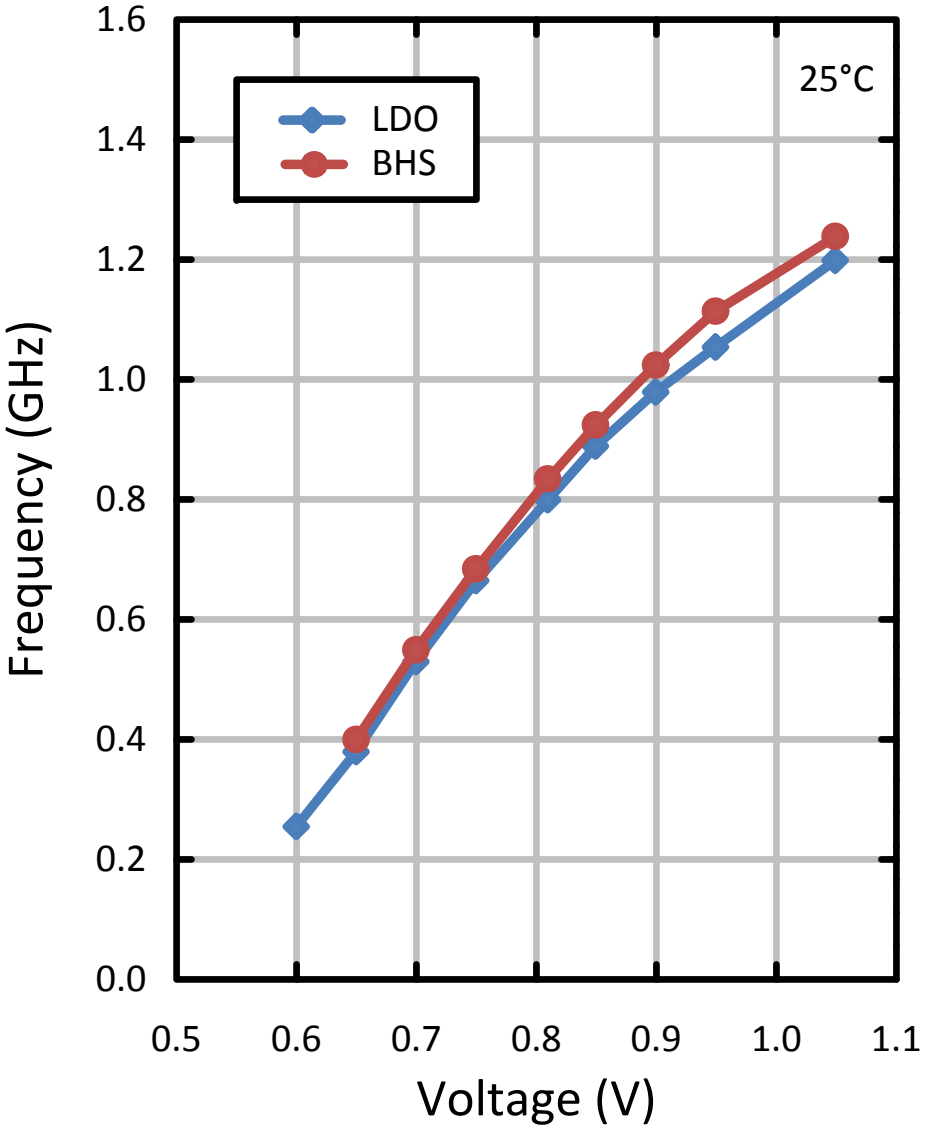
- Introduction
- Core Power Gating and Voltage Scaling
- Low-Power Clocking Techniques
- Asynchronous FIFOs and Arrays
- Core Frequency/Power Data
- Conclusions

Core Leakage

- Performed Monte Carlo simulations to calculate leakage derates for all device types.
- Used cells with different channel lengths and Vt's to minimize leakage on non-critical paths.
  - Considered timing across a wide range of voltages.
- Measured 4.08mW at 0.90V, 25°C.
  - Average of 5 typical parts.
  - From simulations, estimating 10µW after power gating with the BHS.

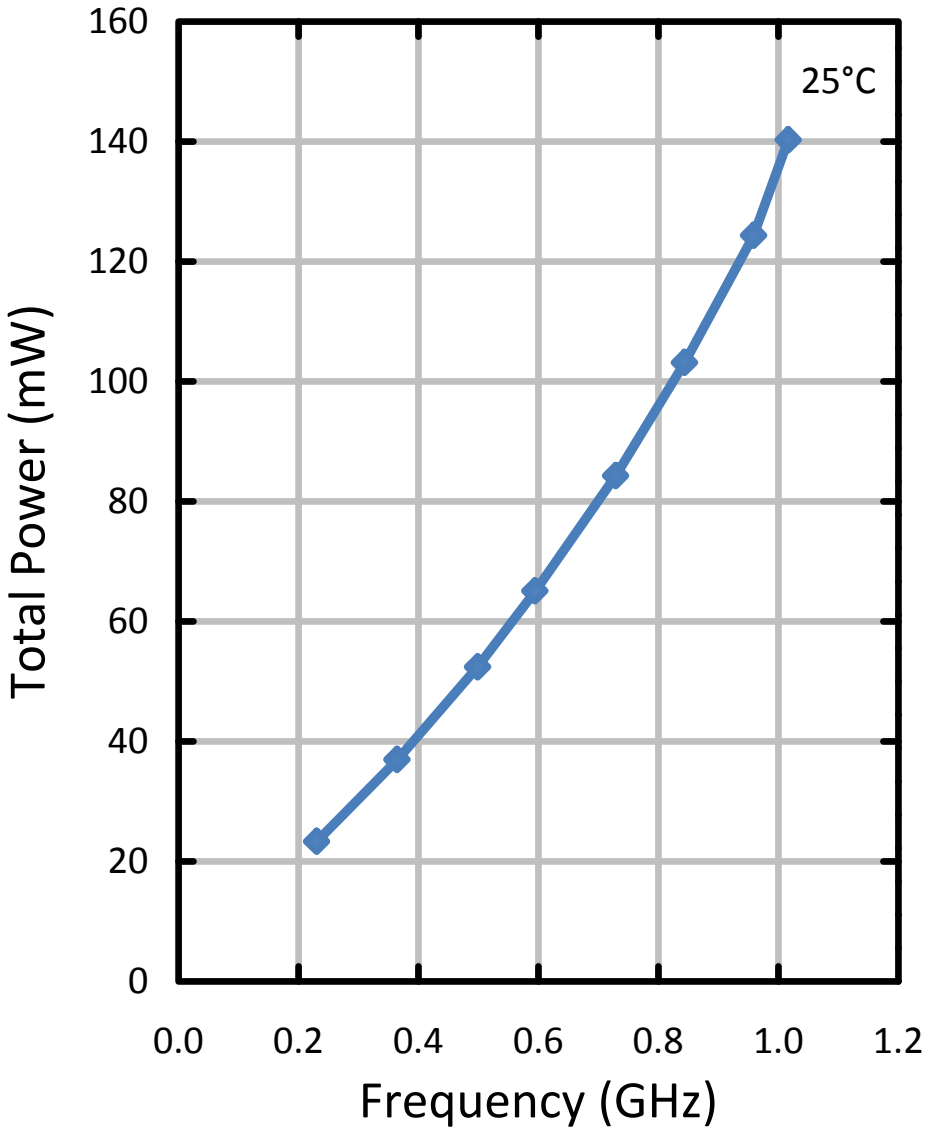
Silicon Results: LDO Versus BHS

- With the LDO, VDDCX is fixed at 1.15V.
- VDDCX cannot go below 0.65V and limits the lowest BHS voltage.



Silicon Results: Total VDDCX Power

- Measurement Conditions:
  - Typical Workload
  - VDDCX = 1.05V
- With VDDCX = 0.70V and VDDQ6 regulated to 0.60V:
  - 13.4mW
  - 58µW/MHz



Outline

- Introduction
- Core Power Gating and Voltage Scaling
- Low-Power Clocking Techniques
- Asynchronous FIFOs and Arrays
- Core Frequency/Power Data
- Conclusions

Comparison

	This work ISSCC 2014	M. Fujigaya et al. ISSCC 2013	Y. Shin et al. ISSCC 2013	S. Jain et al. ISSCC 2012	G. Gammie et al. ISSCC 2011 (1)
Technology	28nm	28nm	28nm	32nm	28nm
I$/D$/L2 (kB)	16/32/256	N/A	32/32/512	8/8/0	32/32/128
On-Chip LDO?	Yes	No	No	No	No
Pulsed Latches?	Yes	N/A	Yes	No	No
Voltage Range	0.60V to 1.05V	1.10V	N/A	0.28V to 1.20V	0.60V to 1.00V
Fmax	1200MHz	1500MHz	1800MHz	915MHz	331MHz
Power at Fmax	197mW	800mW	N/A	737mW	145mW
Best Efficiency	58µW/MHz	533µW/MHz	N/A	170µW/MHz	114µW/MHz
Leakage at 0.90V	4.08mW	N/A	N/A	38mW	20mW

(1) Caches on.

The achieved energy per cycle is 2 or 3 times lower than for the cores optimized for near-threshold computing.

Conclusions

- Introduced a 28nm Hexagon™ VLIW DSP optimized for multimedia and modem applications in a heterogeneous computing environment.
- Supports fine-grain dynamic and leakage power management to provide best-in-class power-efficient performance.
- Unclear that near-threshold computing automatically means better power efficiency.

A 28nm HPM Heterogeneous Multi-Core Mobile Application Processor with 2GHz Cores and Low-Power 1GHz Cores

Mitsuhiko Igarashi[1], Toshifumi Uemura[1], Ryo Mori[1], Noriaki Maeda[1], Hiroshi Kishibe[1], Midori Nagayama[1], Masaaki Taniguchi[1], Kohei Wakahara[1], Toshiharu Saito[1], Masaki Fujigaya[1], Kazuki Fukuoka[1], Koji Nii[1], Takeshi Kataoka[1], Toshihiro Hattori[1]

[1] Renesas Electronics, Tokyo, Japan

mitsuhiko.igarashi.xv@renesas.com

Outline

- Background
- Chip feature
- Heterogeneous architecture and thermal control technique
- Power management technique
  - ✓ Clock tree structure
  - ✓ SRAM macro optimization
  - ✓ Modified adaptive voltage scaling
  - ✓ Real time power saver
- Conclusion

Background

- **Requirement on mobile/car infotainment SoC**
  - Rich user experience and rich entertainment
    - ✓ High peak performance
    - ✓ Continuous high average performance

- **Constraint of mobile devices**
  - Small thermal power budget
    - ✓ Small form factor, no expensive cooling system
  - Battery power operation

Our approach

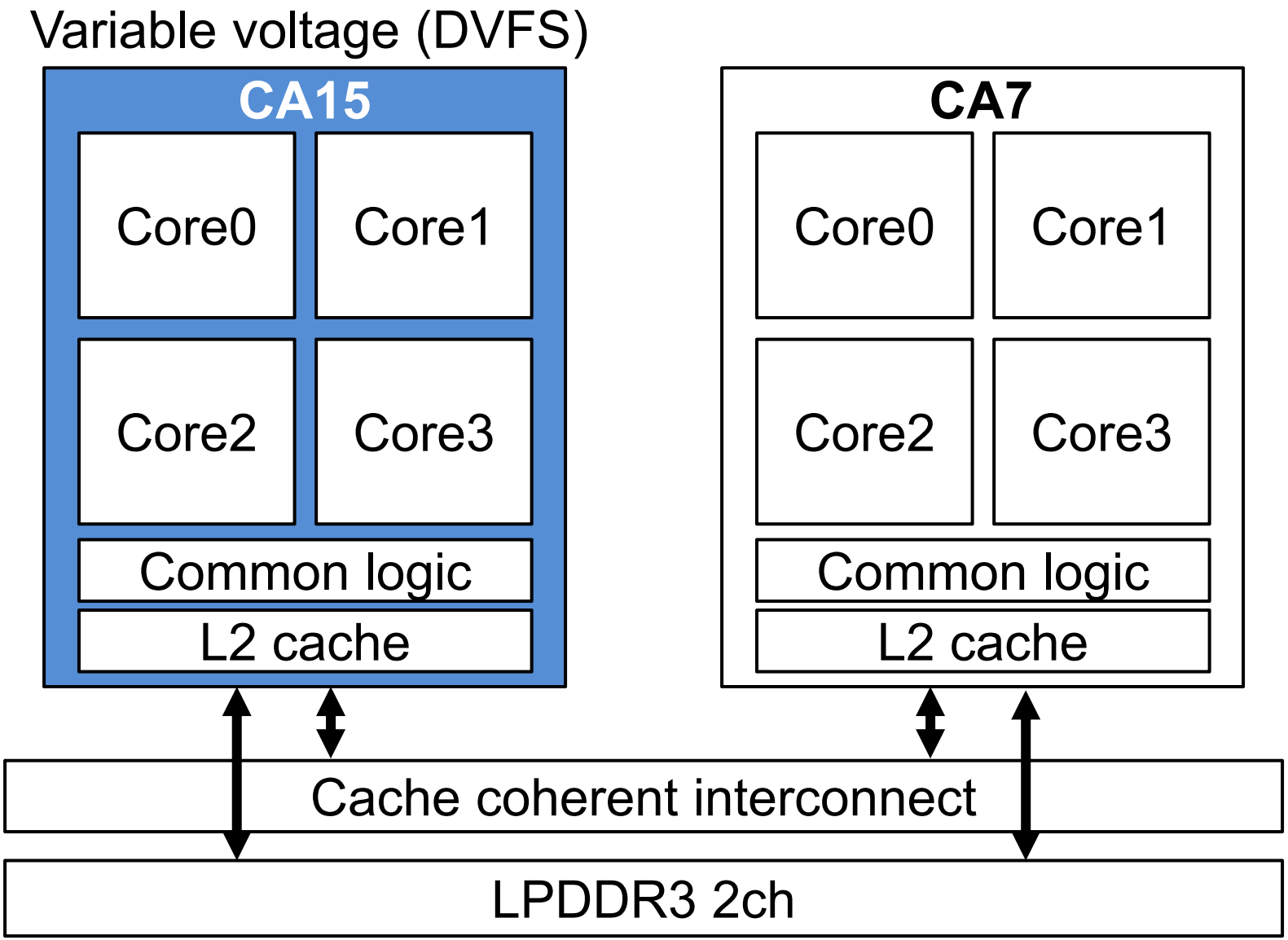
- Heterogeneous octa-core application processor
  - ✓ High-performance 2 GHz quad cores
  - ✓ Simultaneous 8 cores operation
  - ➔Enable high peak performance

- Heterogeneous architecture and several power management techniques
  - ➔Continuous high average performance on small thermal power budget of mobile devices

CPU Comparison

	Renesas's previous work (M. Fujigaya, ISSCC2013)	This work
Process	28nm HKMG Low-leakage CMOS bulk process	28nm HKMG High-Performance for Mobile CMOS bulk process
CPU core	Cortex-A9 (CA9) 1.5 GHz	Cortex-A15 (CA15) 2 GHz Cortex-A7 (CA7) 1 GHz
# core	2	8
DMIPS	7500	35600
Memory	LPDDR2-1066 1ch	LPDDR3-1600 2ch
Power (measured)	0.8 W (CA9 x 2)	6.0 W (total) 5.4 W (CA15 x 4) 0.6 W (CA7 x 4)
Energy efficiency (DMIPS/mW)	9.4 (CA9 x 2)	5.2 (CA15 x 4) 12.7 (CA7 x 4)

Block diagram

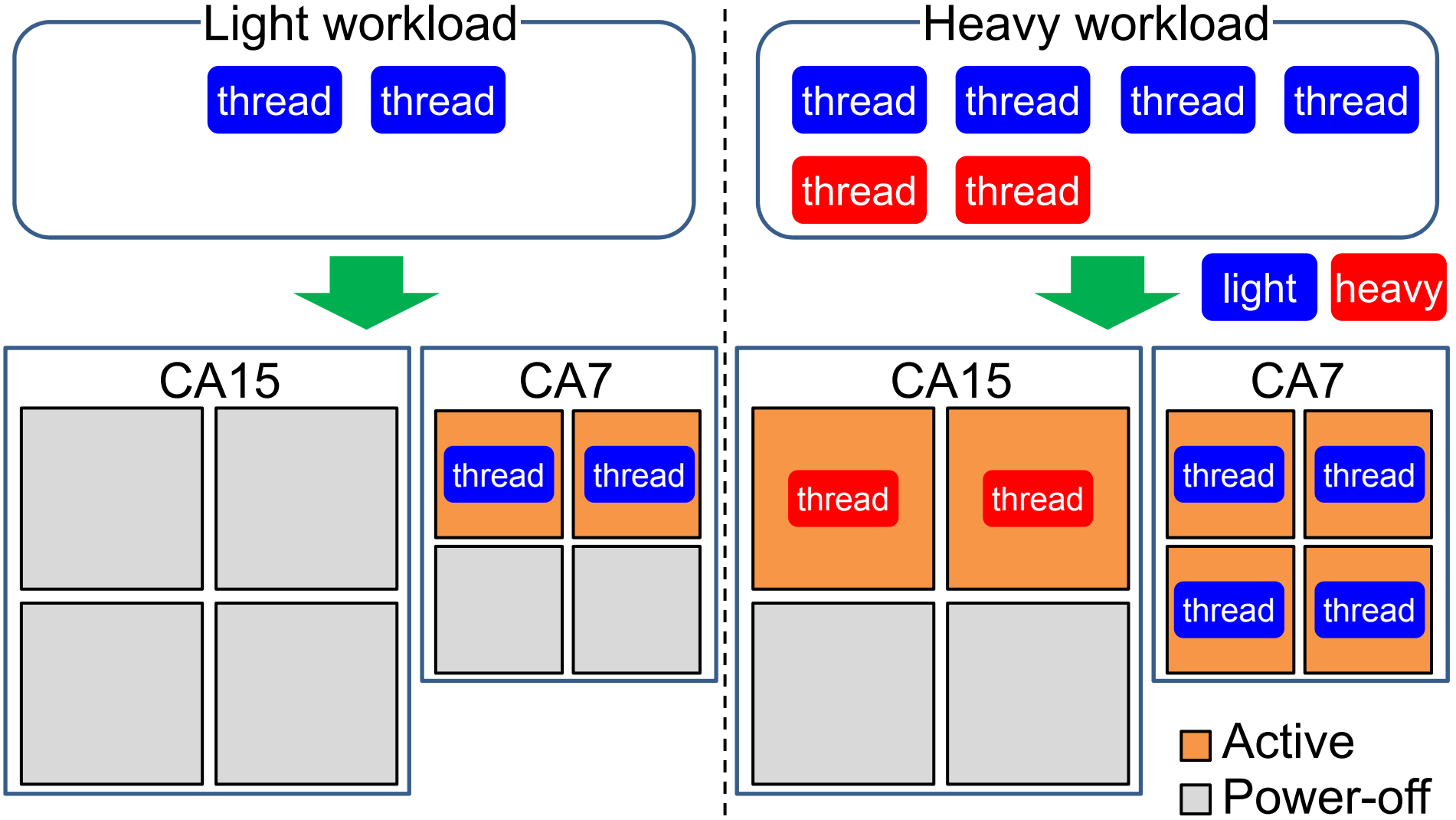


Outline

- Background
- Chip feature
- Heterogeneous architecture and thermal control technique
- Power management technique
  - ✓ Clock tree structure
  - ✓ SRAM macro optimization
  - ✓ Modified adaptive voltage scaling
  - ✓ Real time power saver
- Conclusion

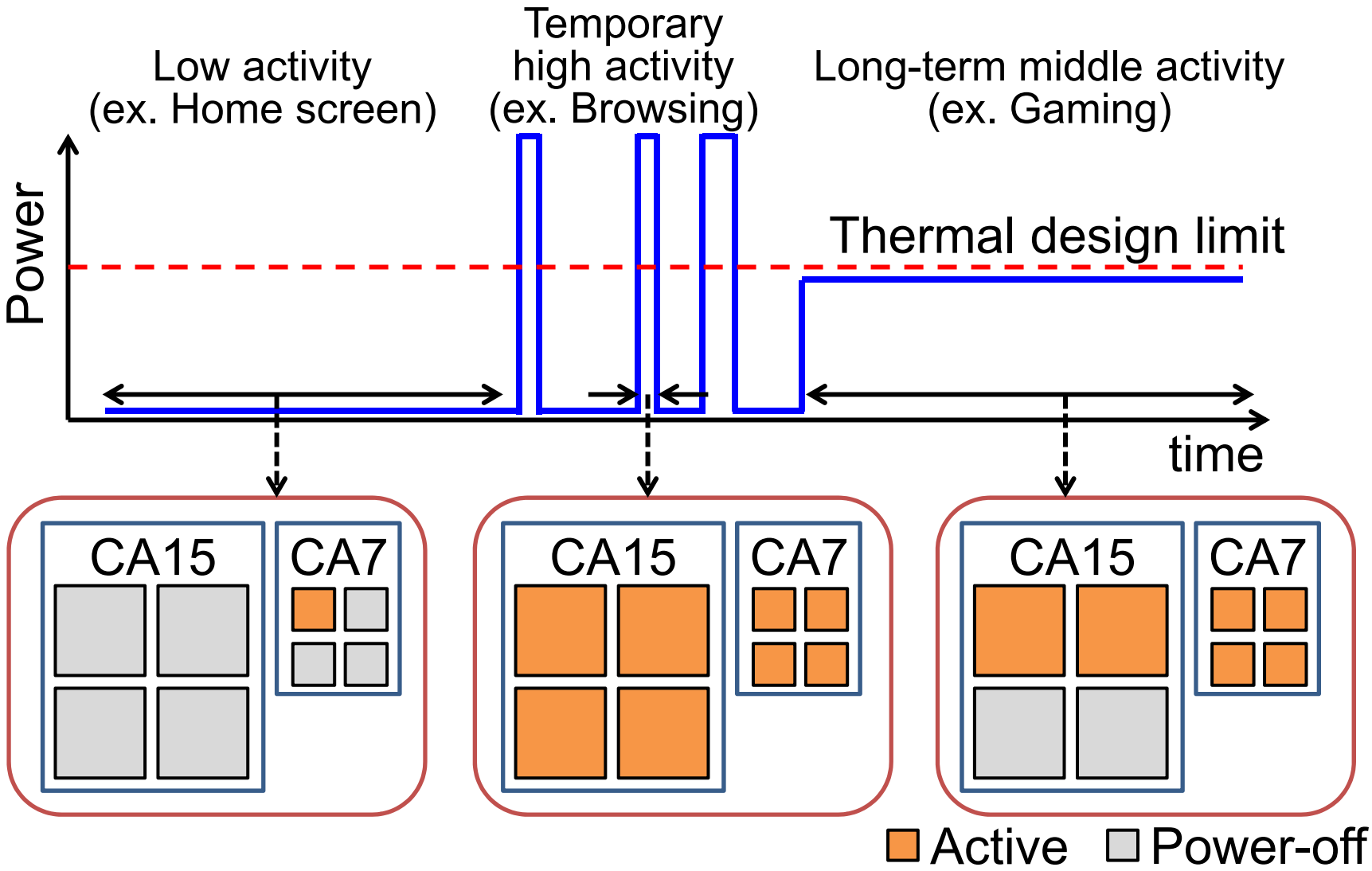
Heterogeneous architecture (1/2)

- Assign tasks to 8 cores depending on each activity



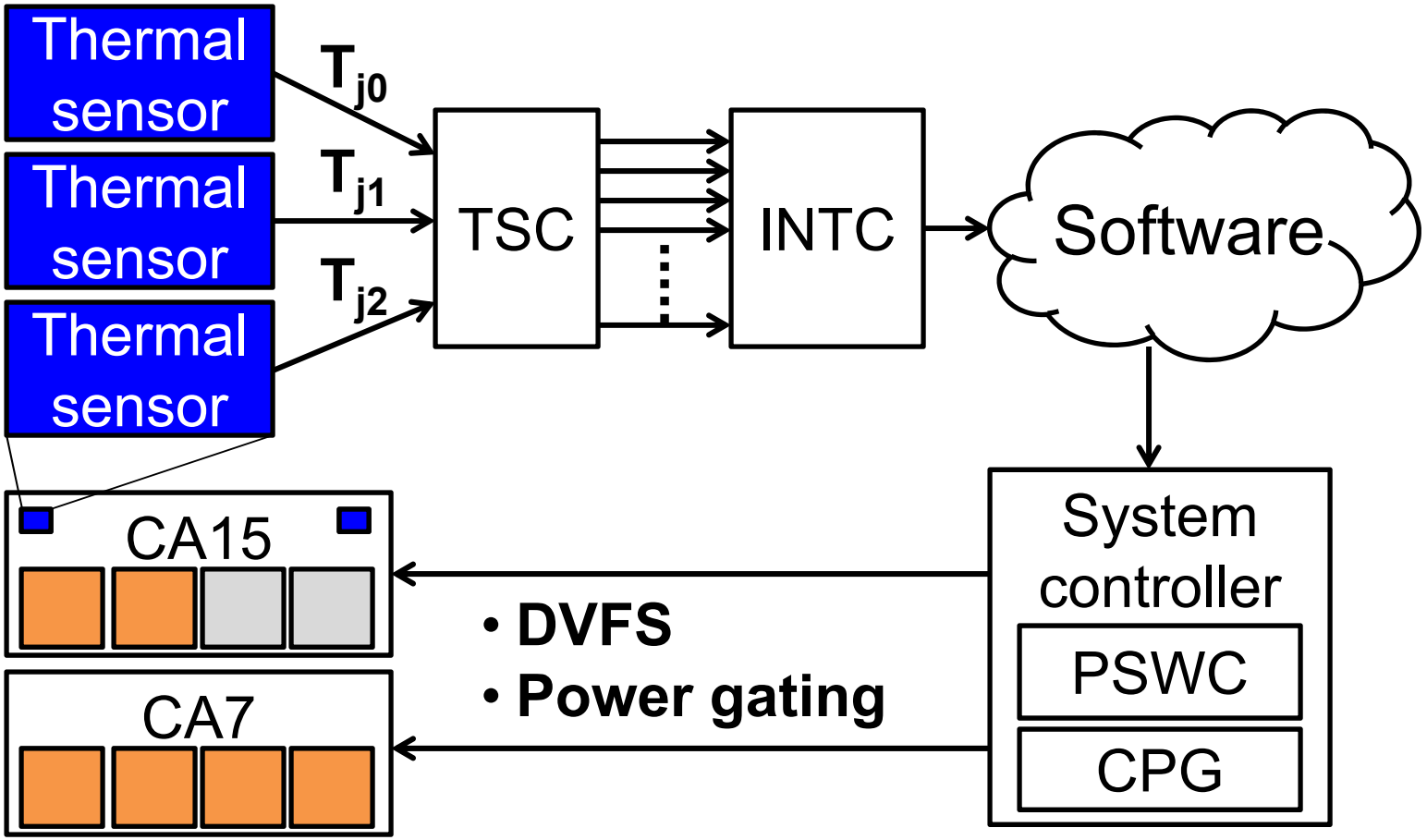
Heterogeneous architecture (2/2)

- Optimal performance and energy-efficiency



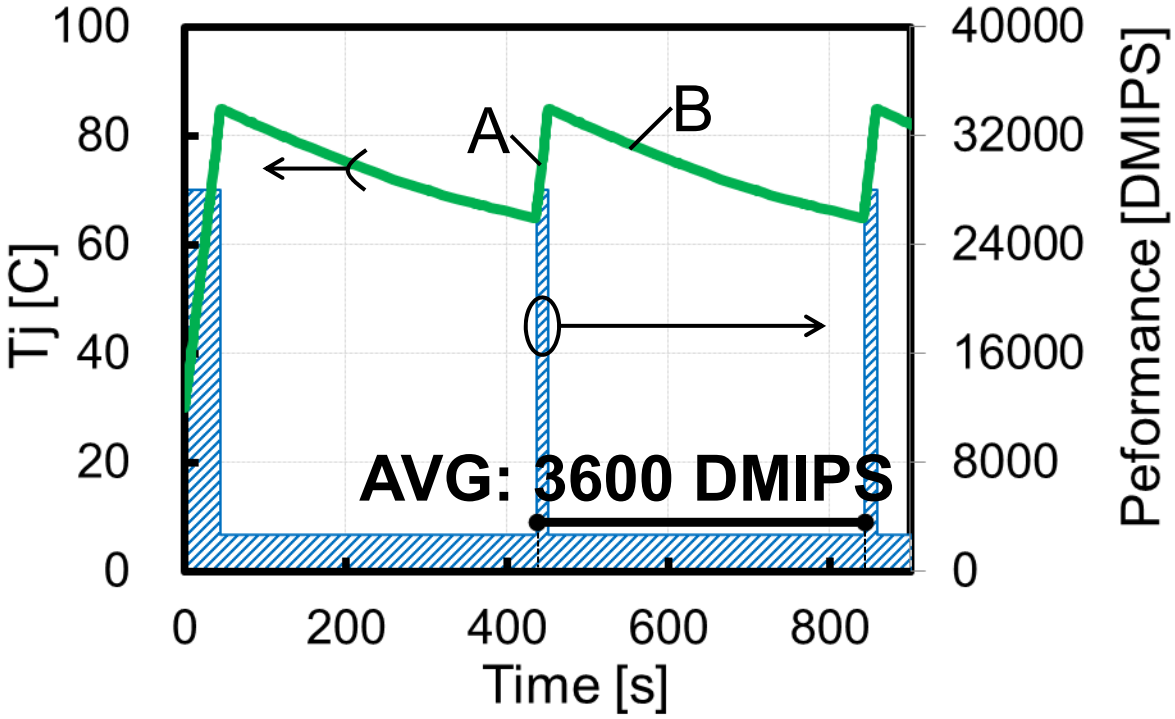
Thermal control technique based on heterogeneous architecture

- Operate energy-efficient CA7 while cooling period



Homogeneous thermal control

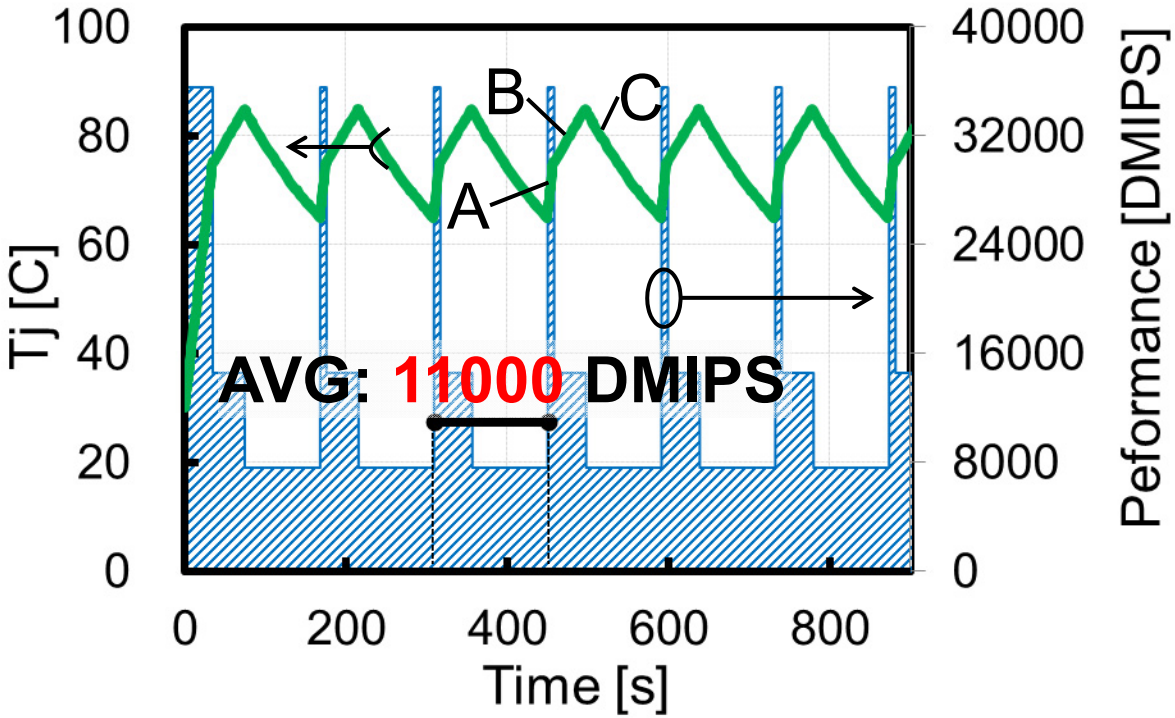
- Significantly decrease of average performance



Mode	CA15				DMIPS
	core0	core1	core2	core3	
A	**2GHz**				**28000**
B	**off**	**off**	**off**	**0.75GHz**	2625

Heterogeneous thermal control

- Satisfy both high peak and average performance



Mode	CA15				CA7				DMIPS
	core0	core1	core2	core3	core1	core0	core0	core0	
A	**2GHz**	**2GHz**	**2GHz**	**2GHz**	**1GHz**	**1GHz**	**1GHz**	**1GHz**	**35600**
B	**off**	**off**	**1GHz**	**1GHz**	**1GHz**	**1GHz**	**1GHz**	**1GHz**	14600
C	**off**	**off**	**off**	**off**	**1GHz**	**1GHz**	**1GHz**	**1GHz**	7600

Outline

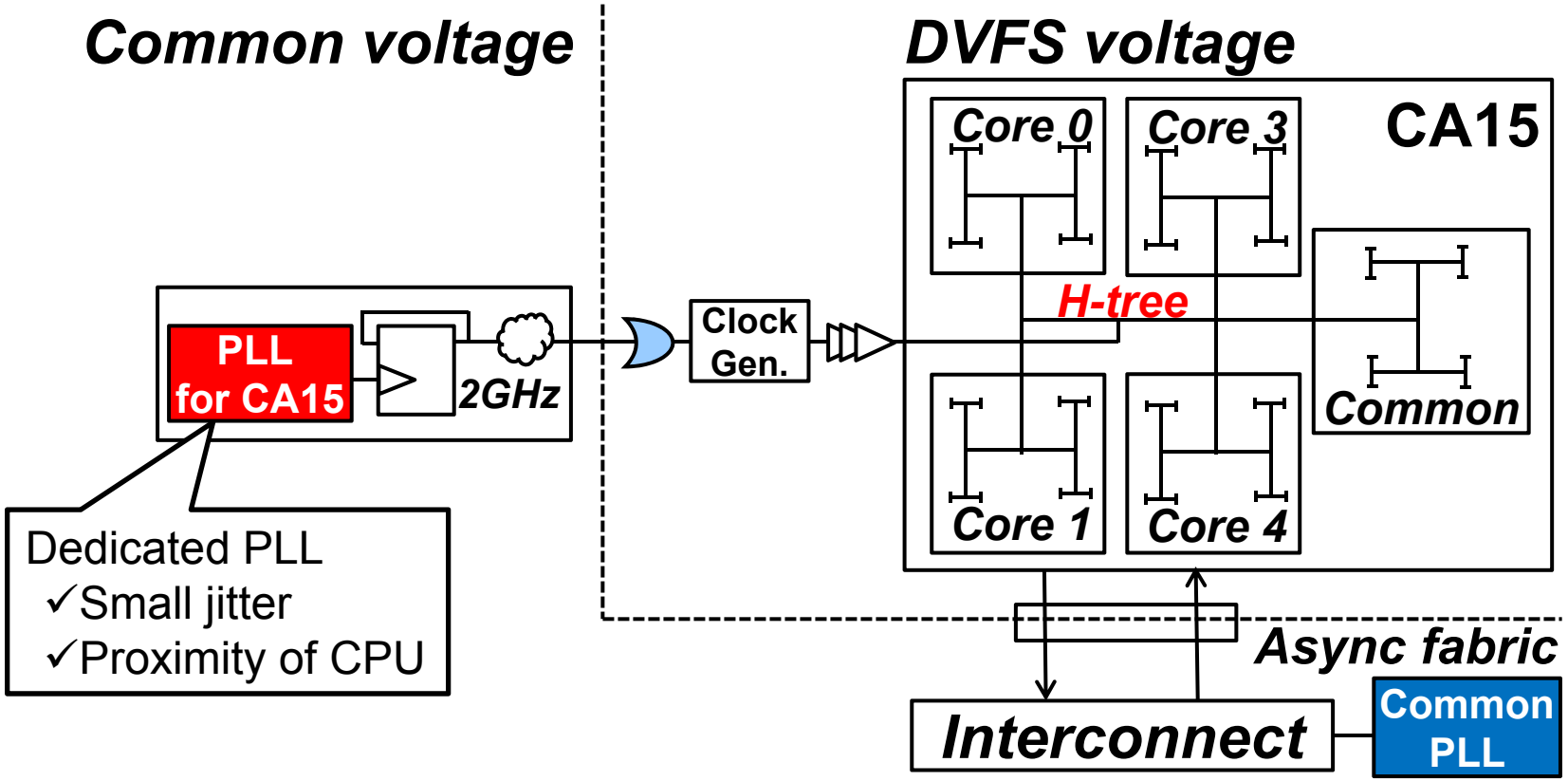
- Background
- Chip feature
- Heterogeneous architecture and thermal control technique
- **Power management technique**
  - ✓ Clock tree structure
  - ✓ SRAM macro optimization
  - ✓ Modified adaptive voltage scaling
  - ✓ Real time power saver
- Conclusion

Low power technique approach of 2 GHz CPU for mobile SoC

- 2 GHz operation and low power
  - ✓ "All H-tree clock structure" and "Dedicated PLL"
  - ✓ Thoroughly optimization of SRAM module by "Multi-Vth" and "Multi-Lg"
  - ✓ Reduction of process variation by modified AVS
- Improve power integrity for lower voltage operation
  - ✓ Reduction of voltage variation by modified AVS
  - ✓ Suppress AC IR drop by Real time power saver

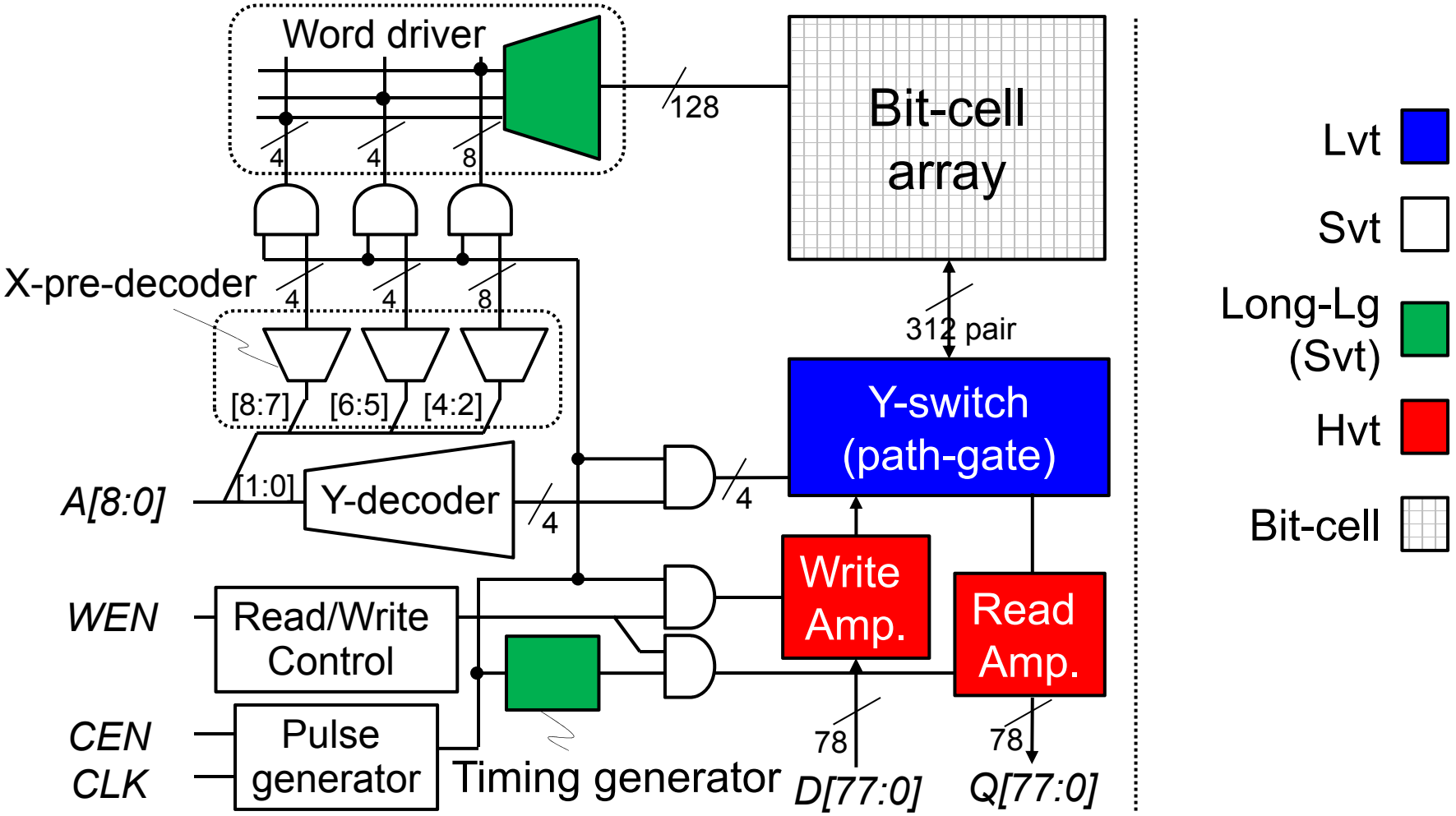
CPU Clock distribution

- All H-tree clock structure can reduce
  - ✓ Wiring capacitance of the clock tree
  - ✓ Drivability of the clock buffer
- Dedicated PLL reduces clock latency and jitter



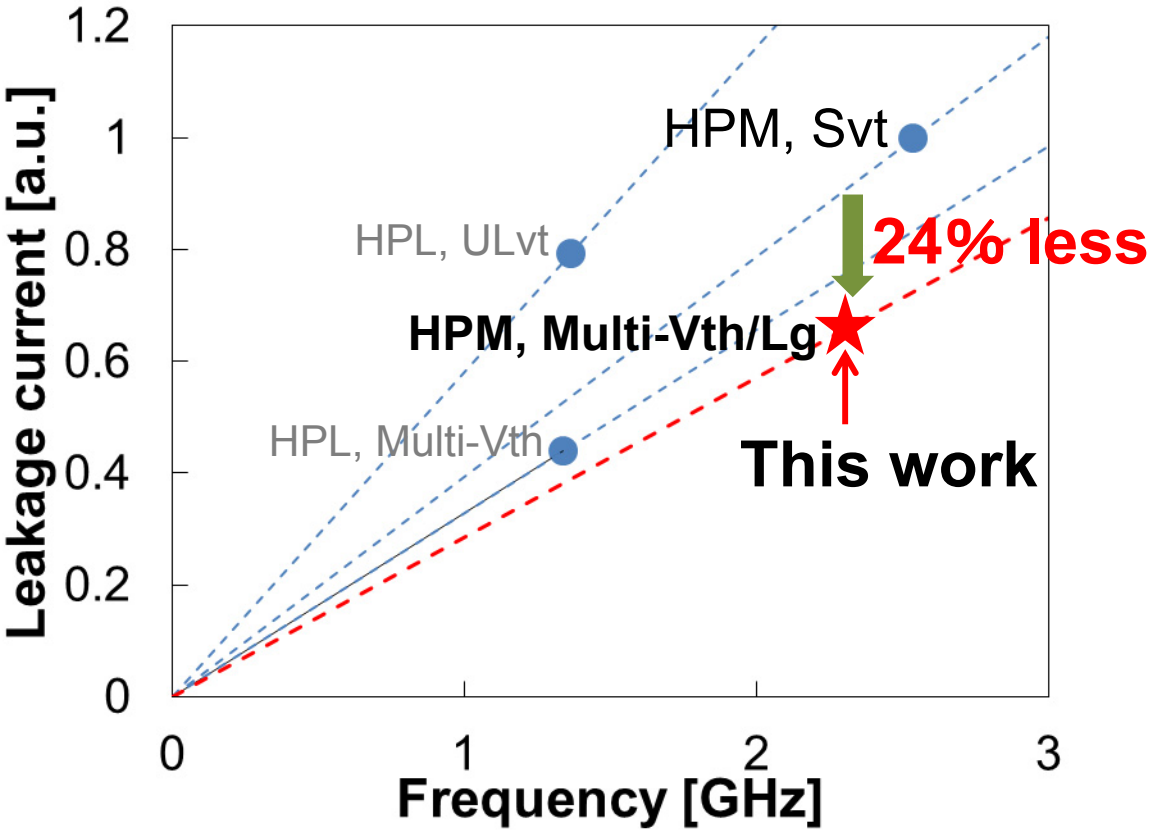
Optimization of SRAM module (1/2)

- Through optimization by multi-Vth and multi-Lg



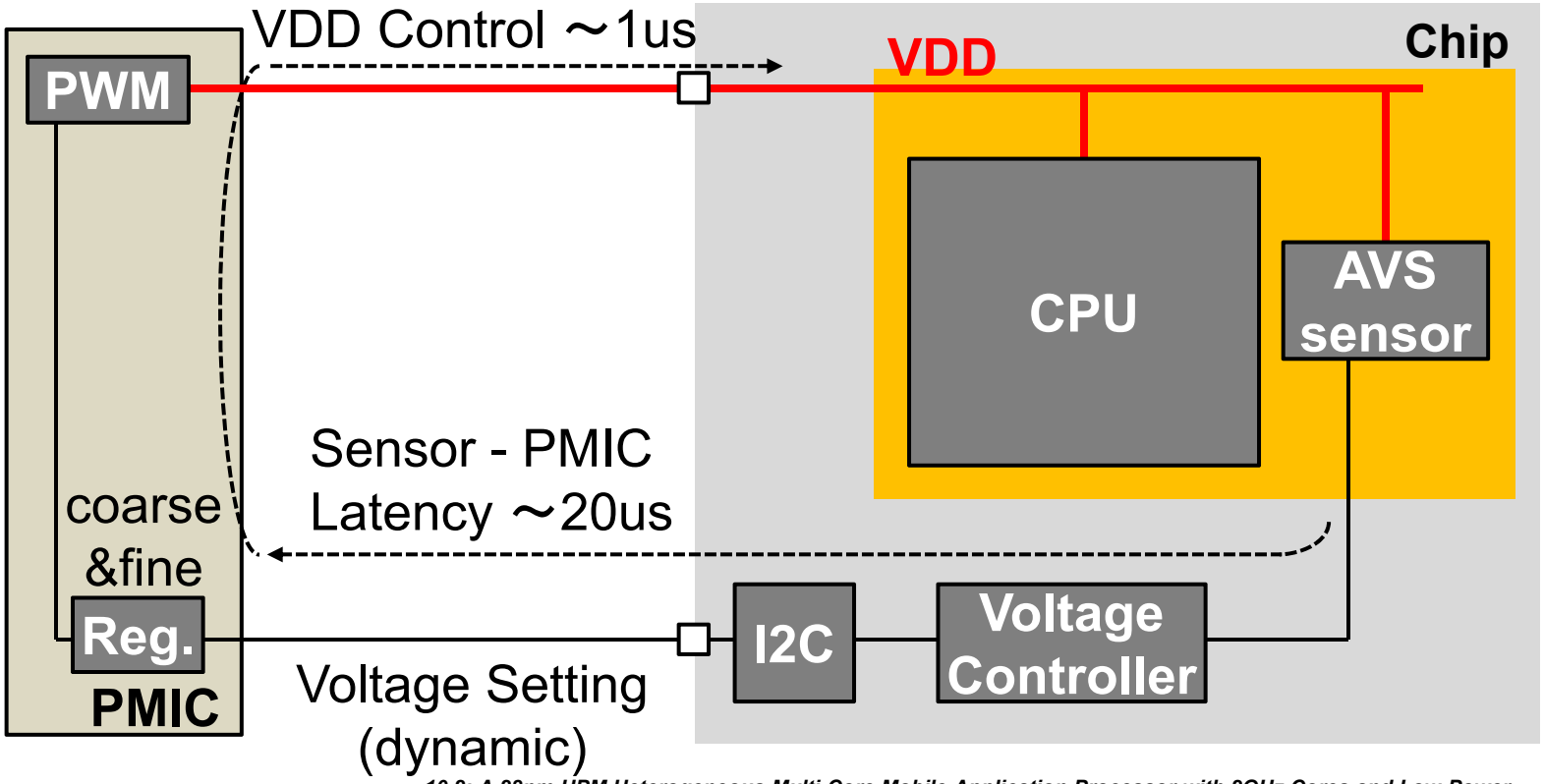
Optimization of SRAM module (2/2)

- 24% leakage current reduction of SRAM module



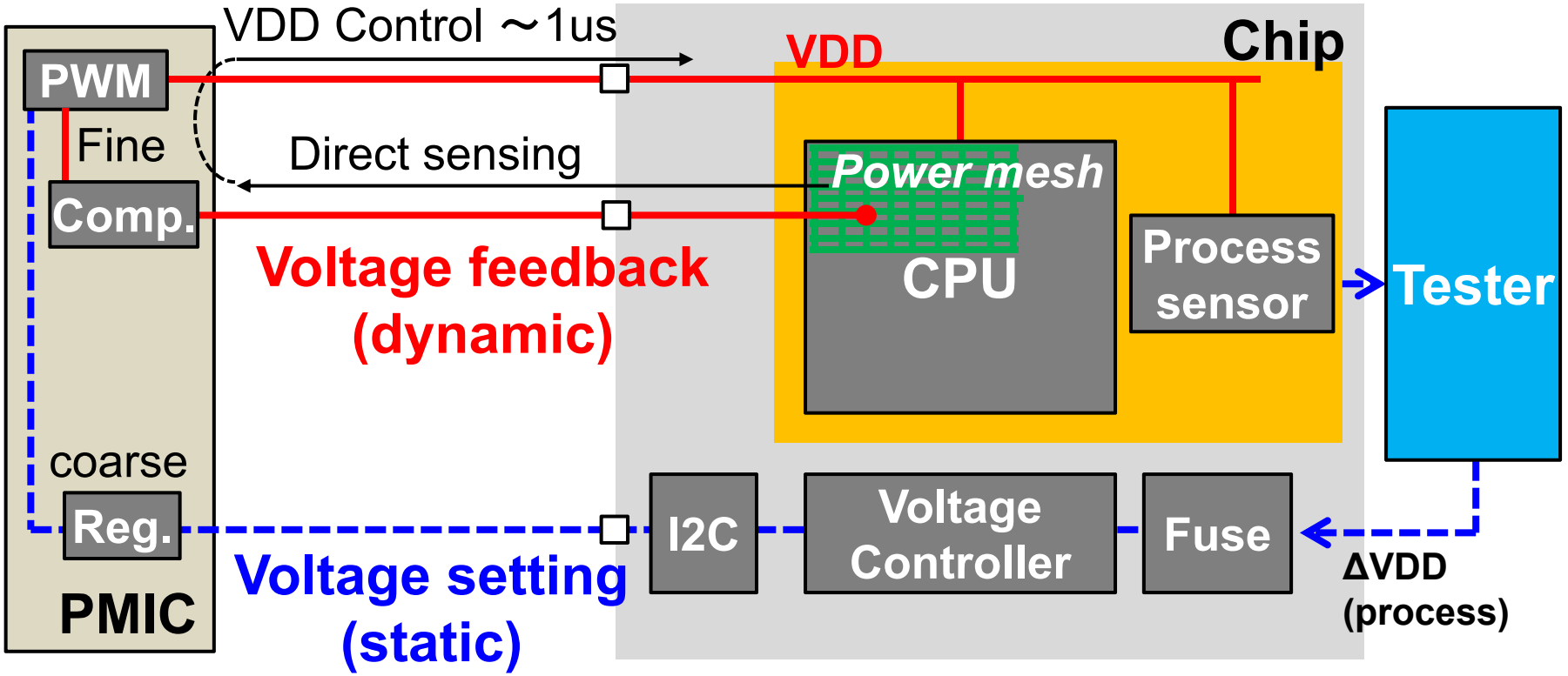
Conventional Adaptive Voltage Scaling

- Problems of conventional AVS
  - ✓ Need voltage guardband for sensor accuracy
  - ✓ Large communication latency b/w chip and PMIC



Modified Adaptive Voltage Scaling

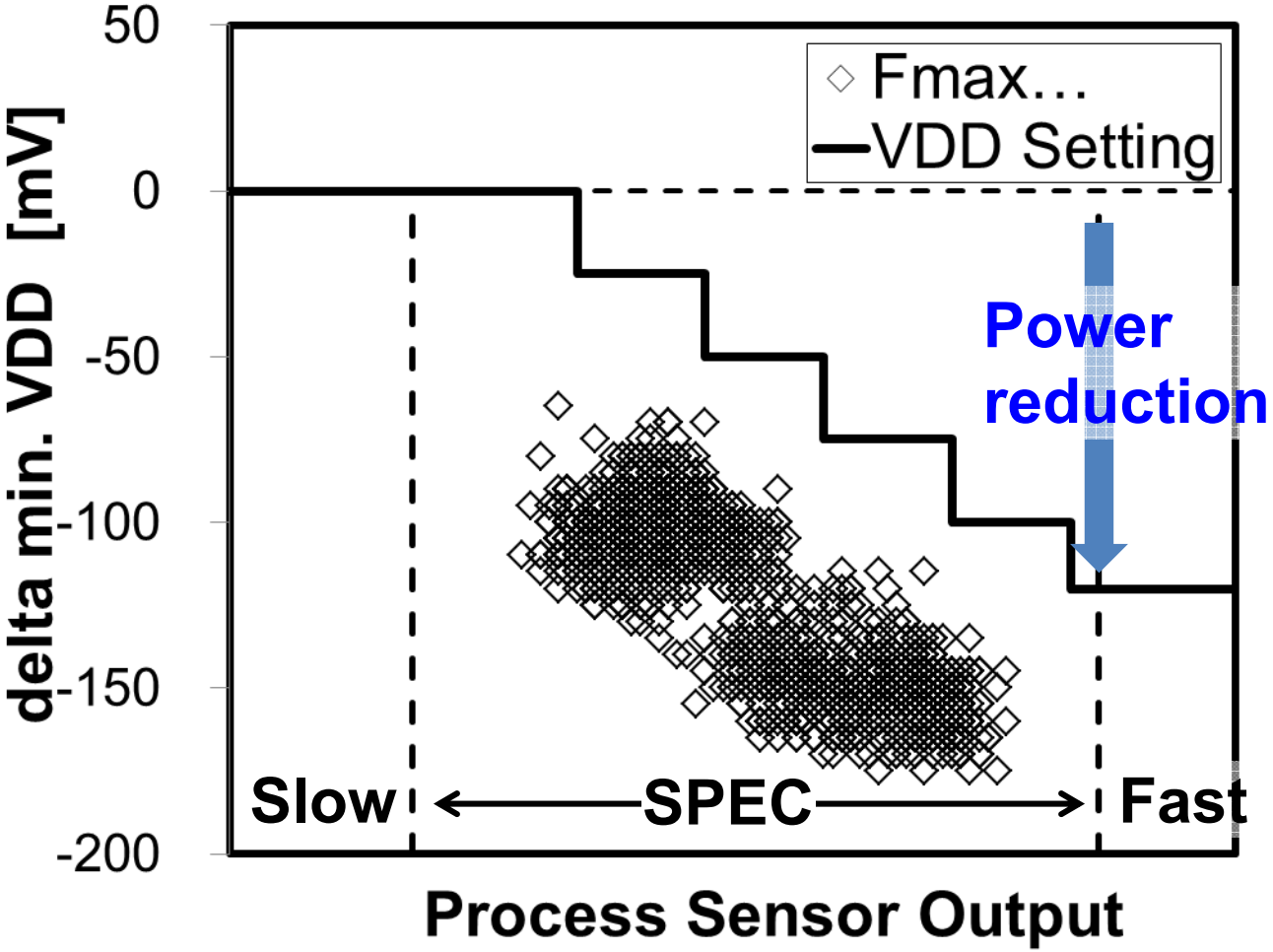
- Static adjustment of process variation at testing
- Improve accuracy and feedback speed by direct sensing of power mesh



Measurement result of AVS (1/2)

- VDD control to reduce process variation

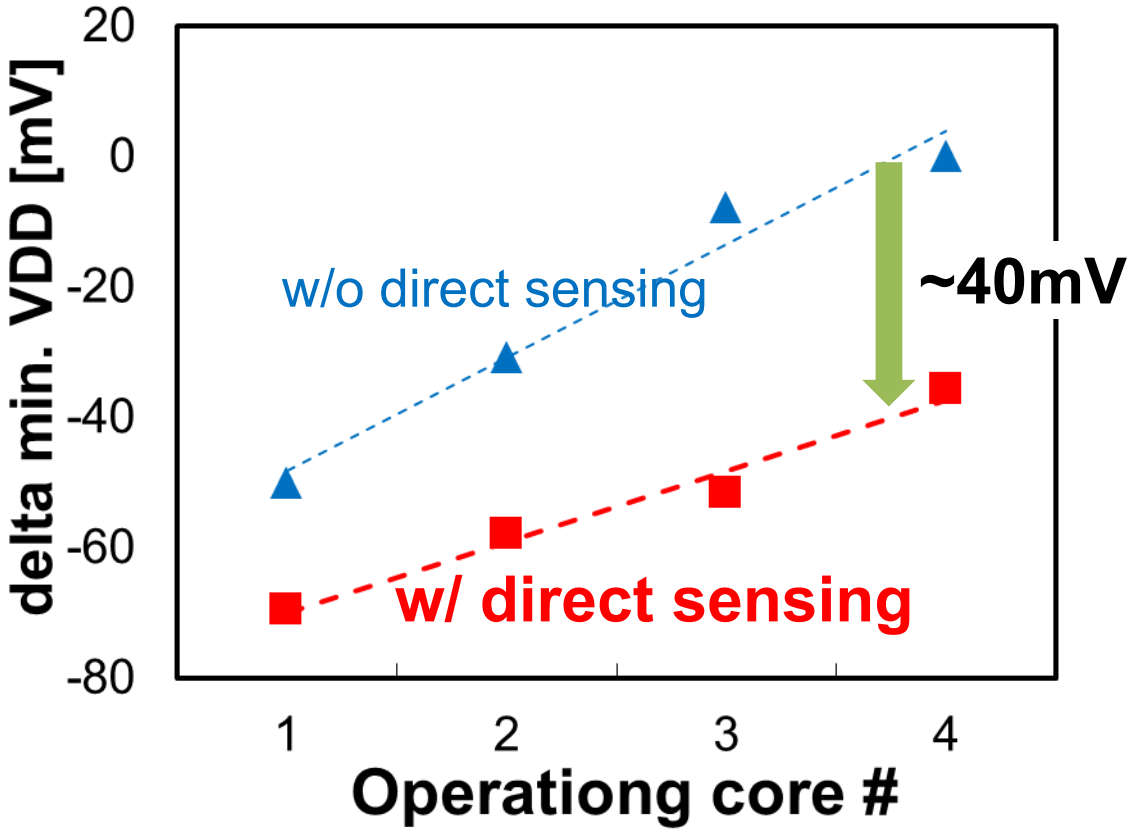
➔ -20% and -29% of dynamic and leakage power



Measurement result of AVS (2/2)

- Direct sensing reduces voltage variation

➔ ~ 40 mV improvement of min. VDD



CPU clock control (1/3)

- Large di/dt at high-performance 2 GHz CPU produces excessive AC voltage drop

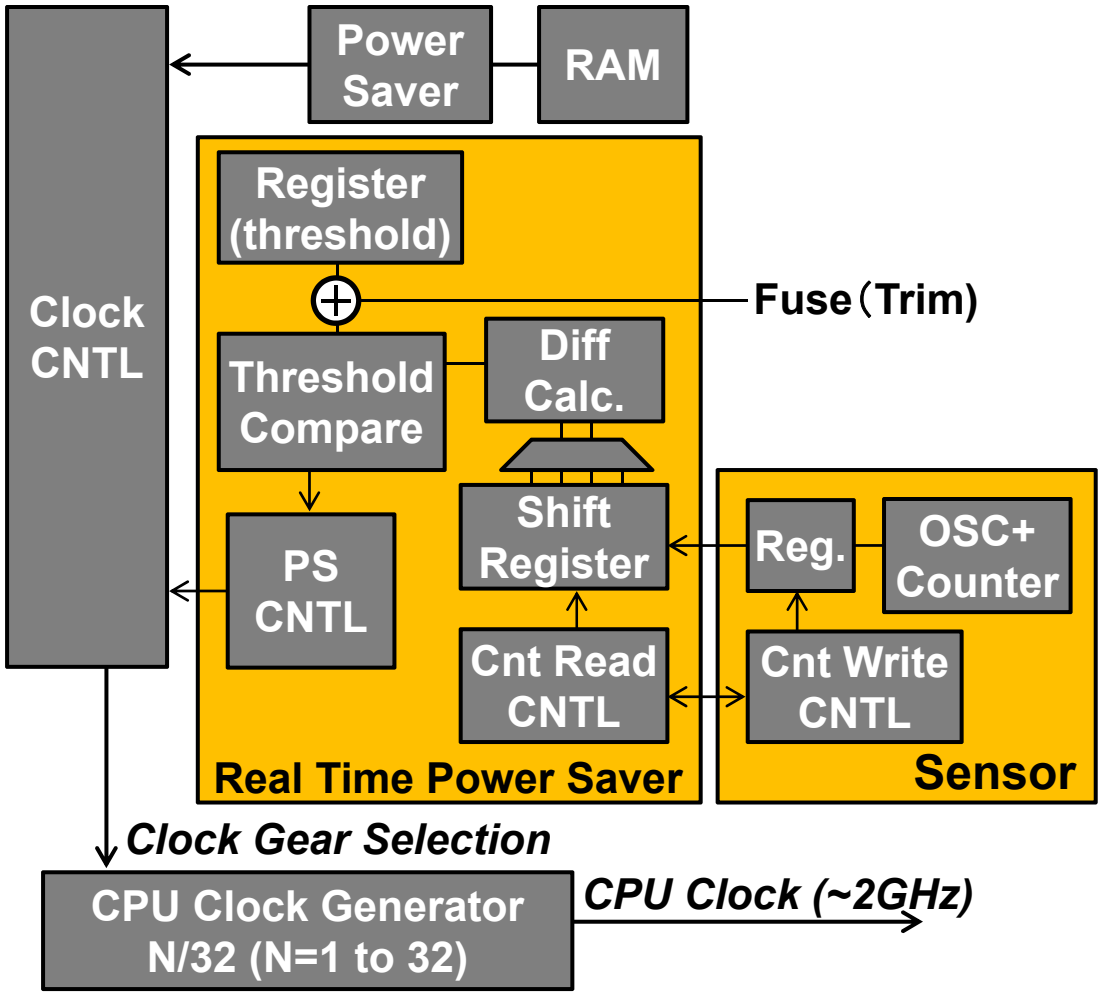
  ➔ Dynamic frequency control technique

- Sampling rate of conventional power saver : ~1us

  ➔ 20x faster sampling rate by on-chip sensor

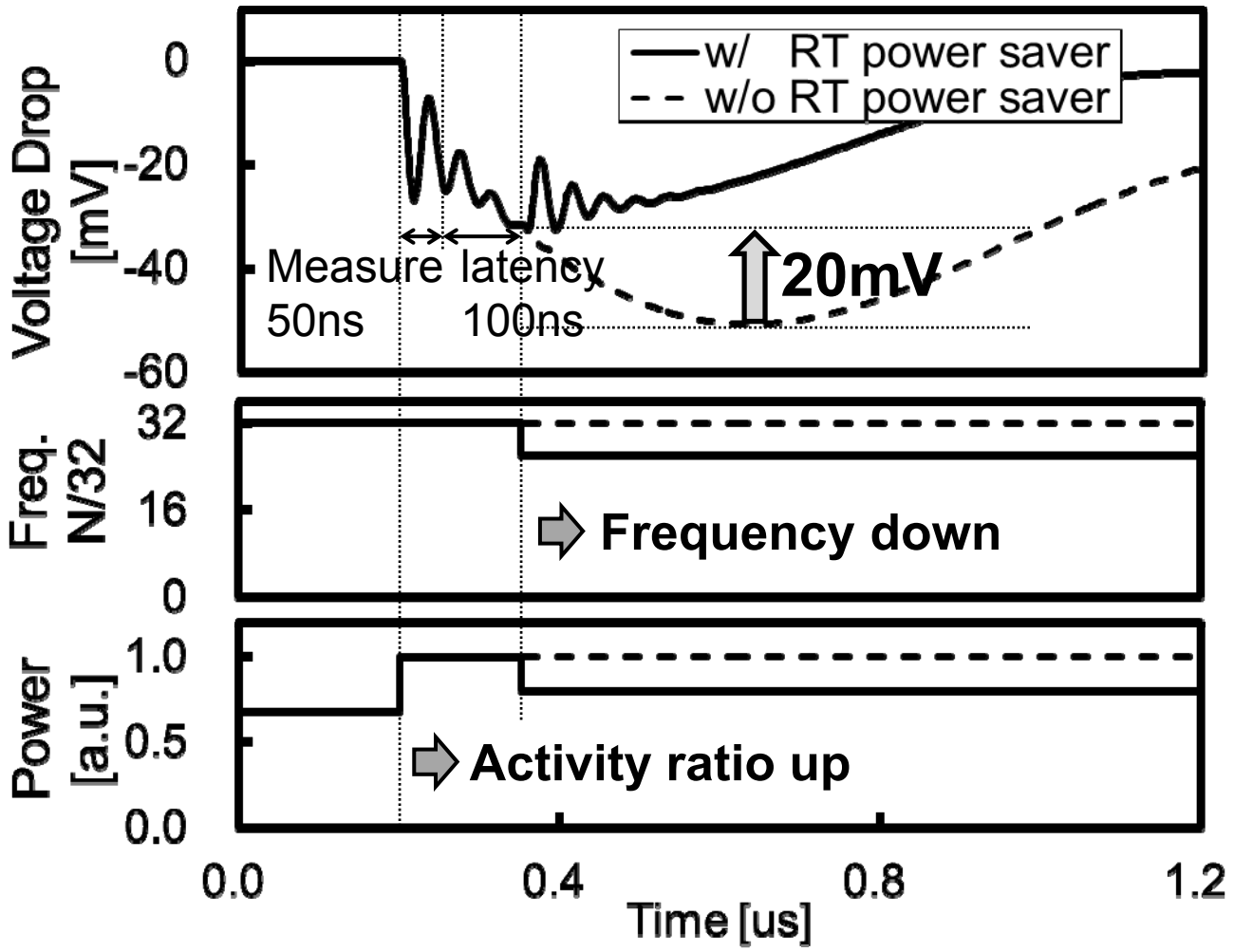
CPU clock control (2/3)

- Monitor the difference of count in the 50ns interval

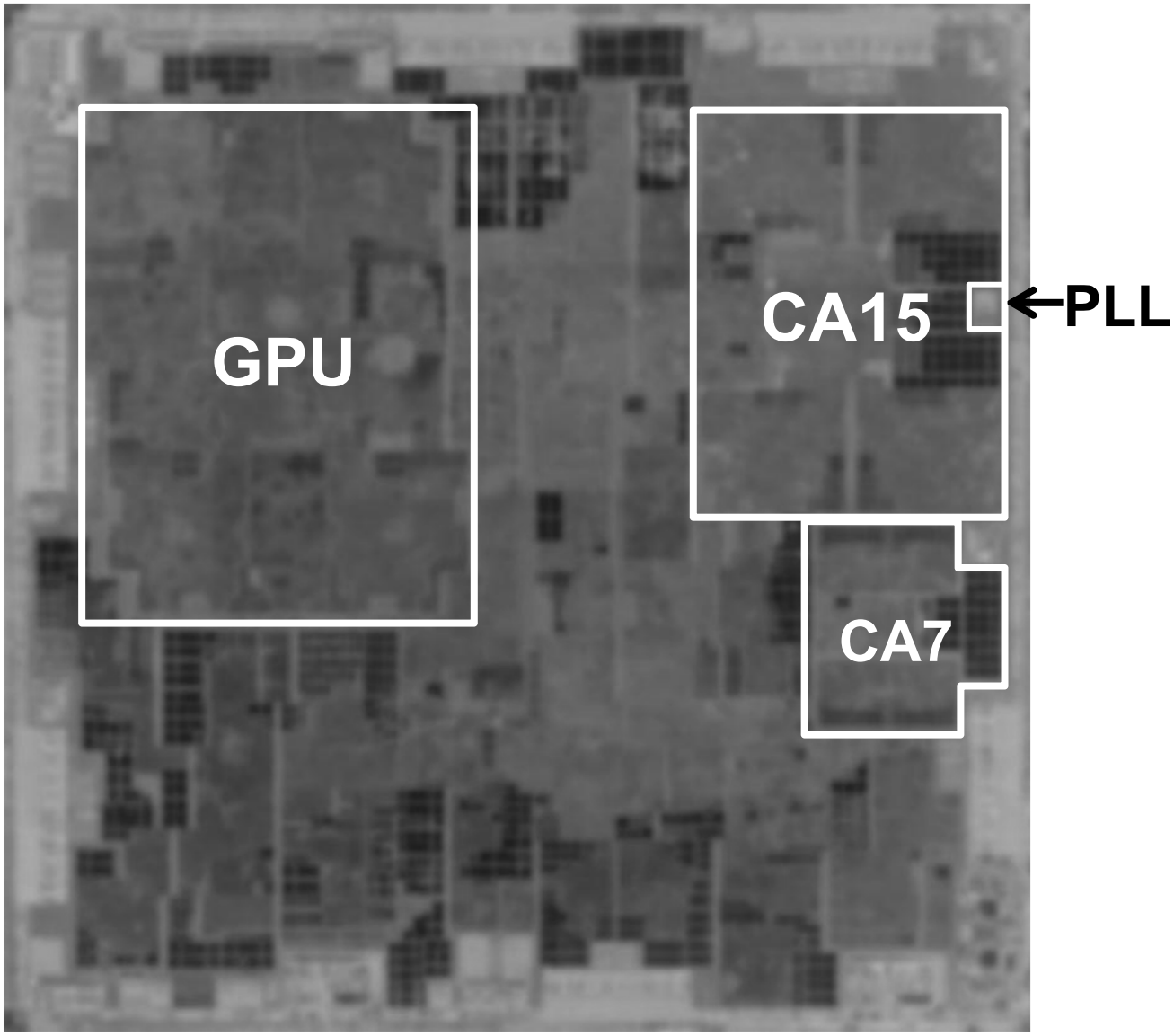


CPU clock control (3/3)

- 20mV improvement of AC Vdrop by simulation



Die micrograph



Conclusion

- Heterogeneous octa-core application processor with high performance 2 GHz CPU
- Heterogeneous architecture and low power enable continuous high average performance in mobile devices
- Several power management technique

1. Low power design of clock tree and SRAM module
2. Modified Adaptive Voltage Scaling technique
   - ✓ 20% and 29% reduction of dynamic and leakage power
   - ✓ 40mV min. VDD improvement
3. Real time power saver w/ 20x fast sampling rate

2014 ISSCC paper 10.3

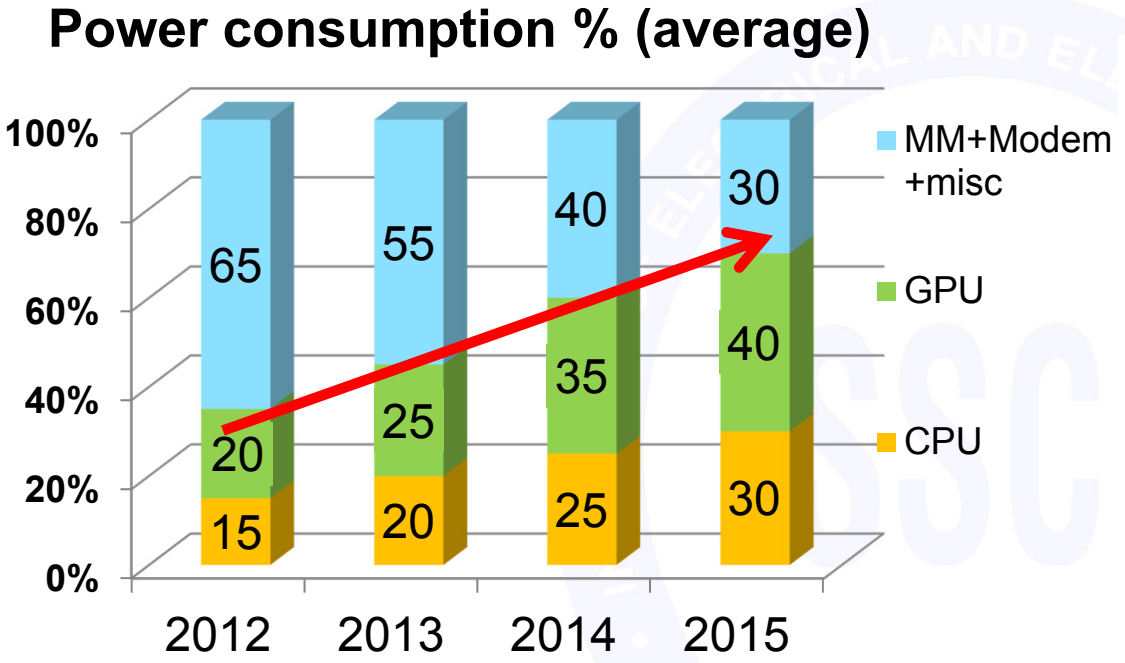
Heterogeneous Multi-Processing Quad-Core CPU and Dual-GPU Design for Optimal Performance, Power, and Thermal Tradeoffs in a 28nm Mobile Application Processor

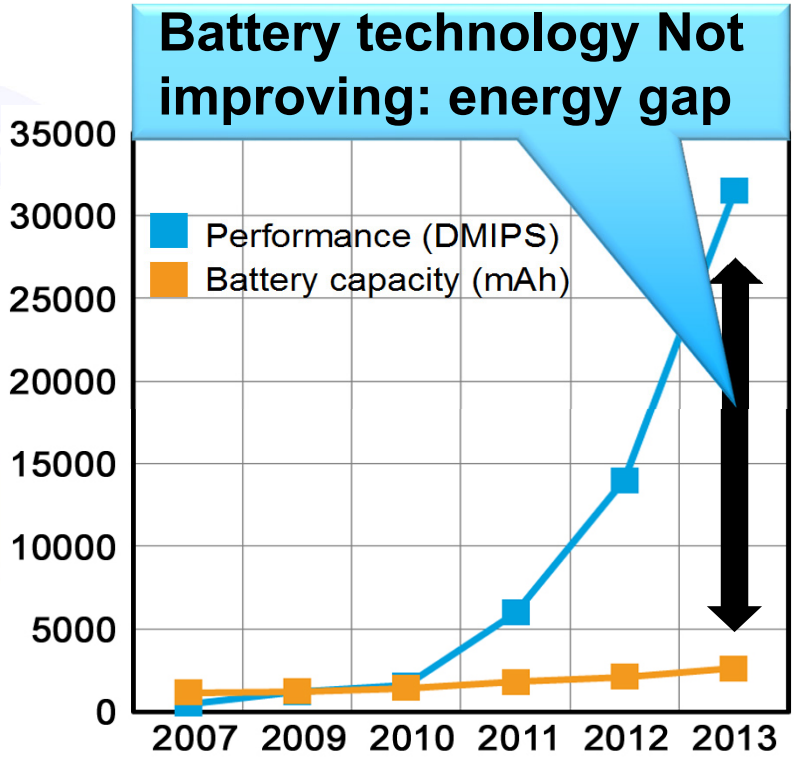
Alice Wang, Tsung-Yao Lin, *Shichin Ouyang*, Wei-Hung Huang, Jidong Wang, Shu-Hsin Chang, Sheng-Ping Chen, Chun-Hsiung Hu, Jim C. Tai, Koan-Sin Tan, Meng-Nan Tsou, Ming-Hsien Lee, Gordon Gammie, Chi-Wei Yang, Chih-Chieh Yang, Yeh-Chi Chou, Shih-Hung Lin, Wuan Kuo, Chi-Jui Chung, Lee-Kee Yong, Chia-Wei Wang, Kin Hooi Dia, Cheng-Hsing Chien, You- Ming Tsao, Nitin Kumar Singh, Rolf Lagerquist, Chih-Cheng Chen, Uming Ko

Outline

- Introduction
- Processor Technology
- Circuit Technology
- System Technology
- Summary

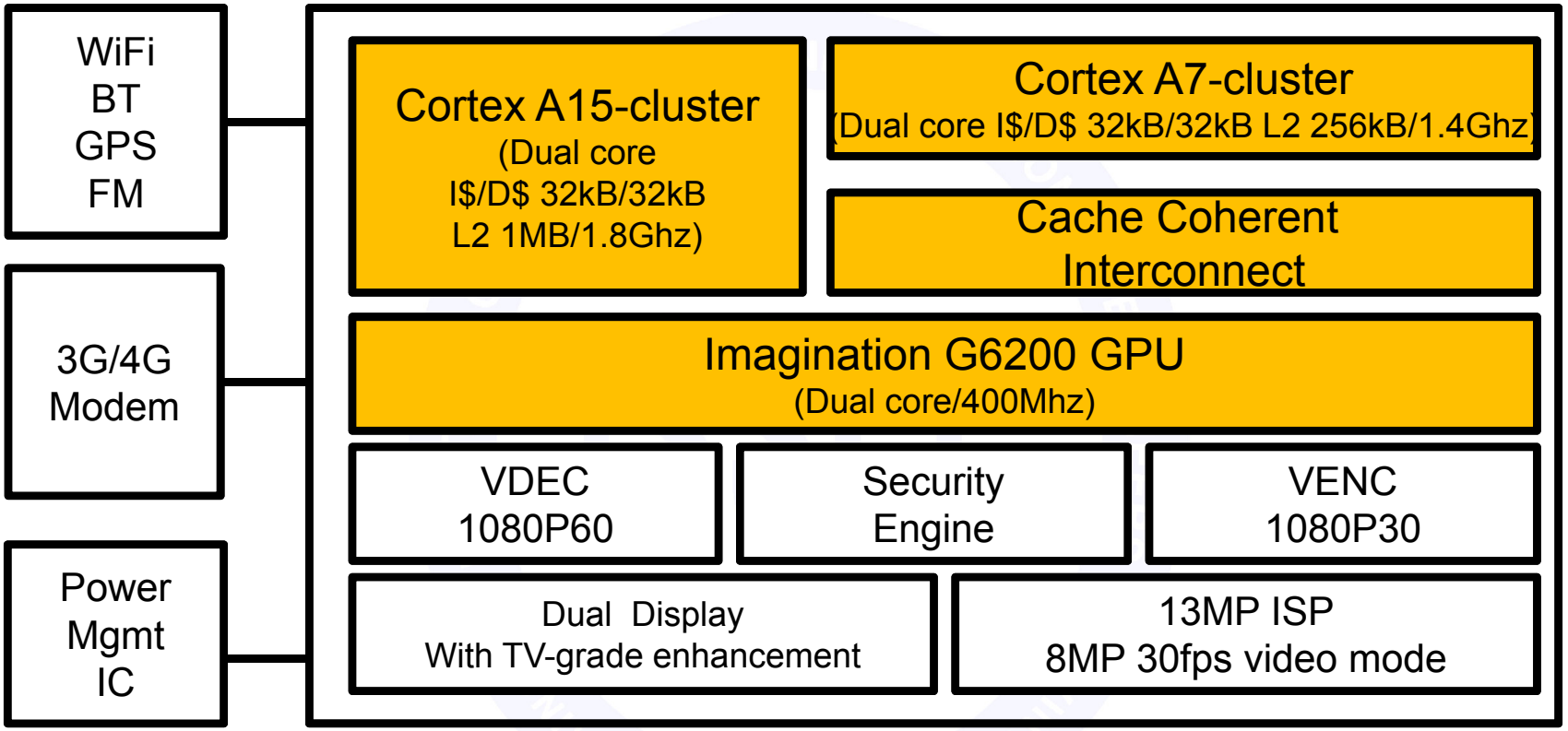
CPU+GPU power trend





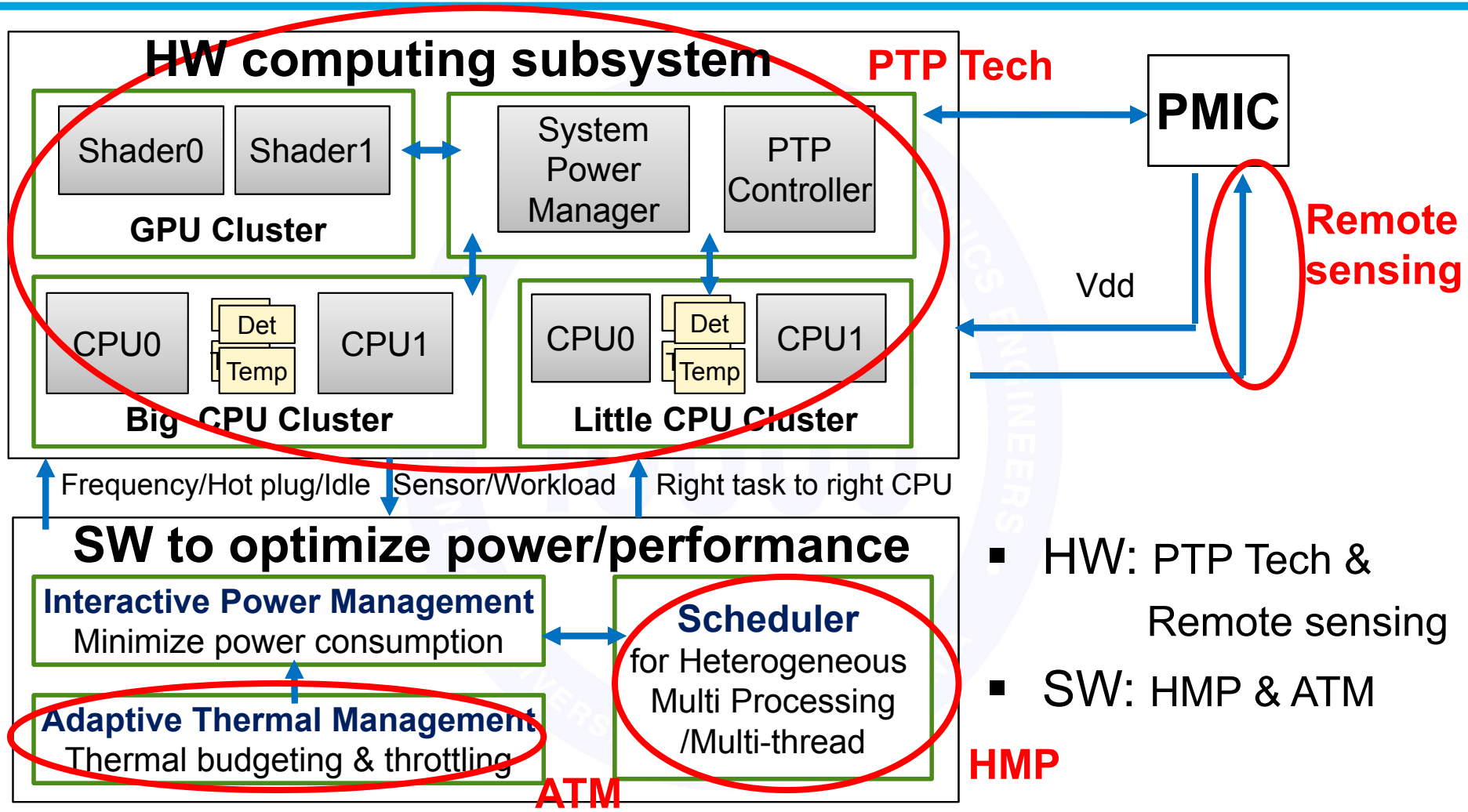
- CPU and GPU power consumption % is growing on mobile SoCs
- CPU+GPU power can together consume up to 90% of total power when handling compute-intensive tasks
- Innovation required to keep within fixed power and thermal budgets while providing high performance to close the energy gap

28nm AP overview



- Heterogeneous ARM-v7A quad-core CPUs and a dual-core GPU
- Challenges: Performance, Thermal, Power
- Solutions: PTP Tech, Remote Sensing, LP SRAM, HS CG, HMP, ATM

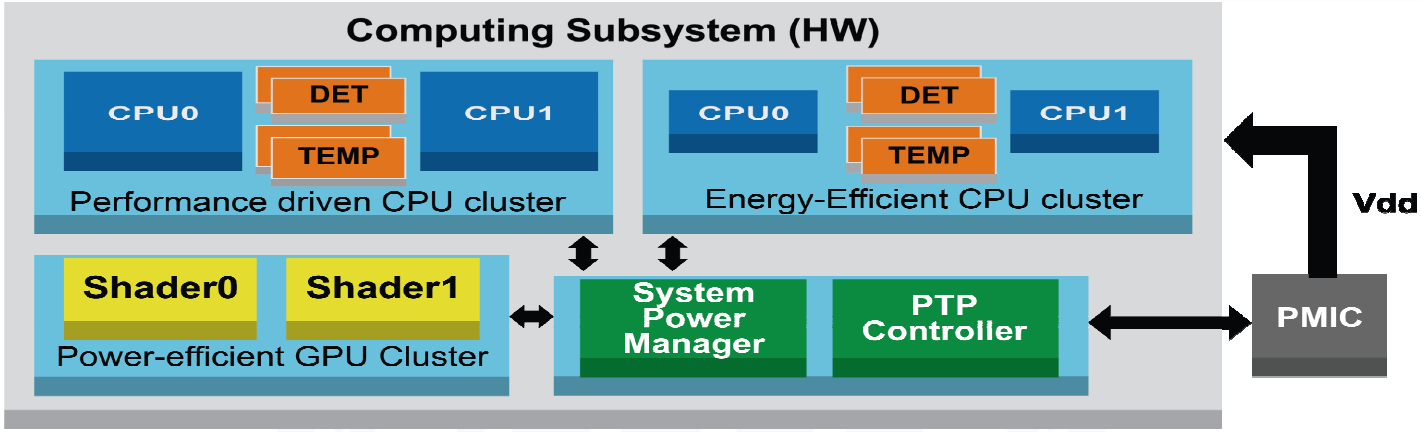
Low power system architecture

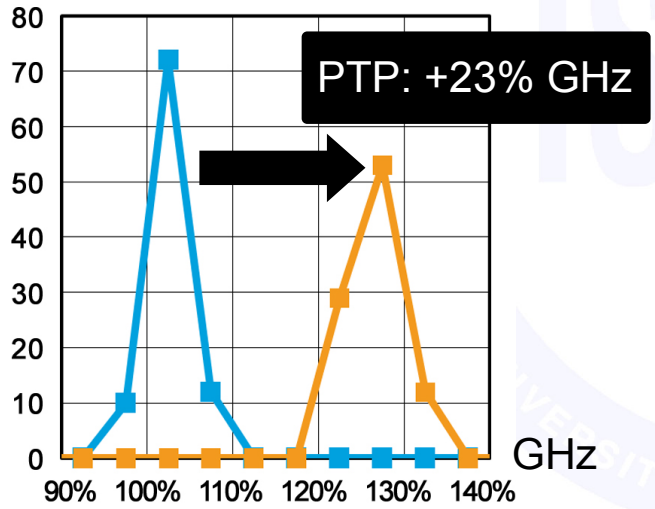


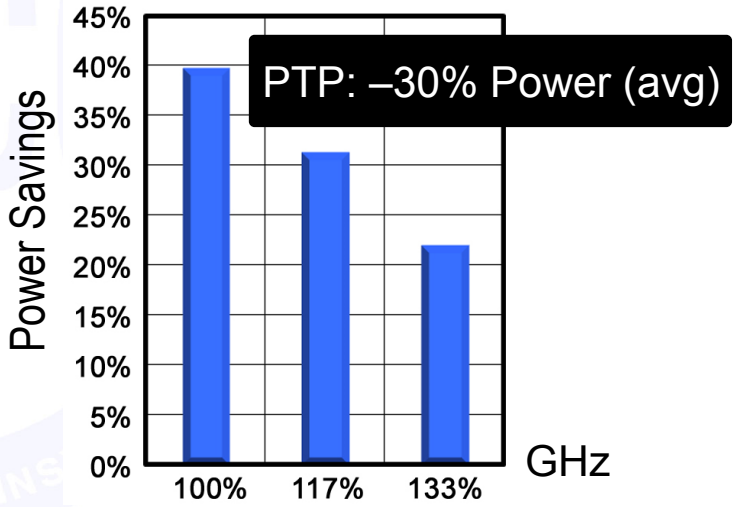
- HW: PTP Tech & Remote sensing
- SW: HMP & ATM

- HW & SW co-optimization for thermal, power and performance

Performance, Thermal & Power (PTP) Technology

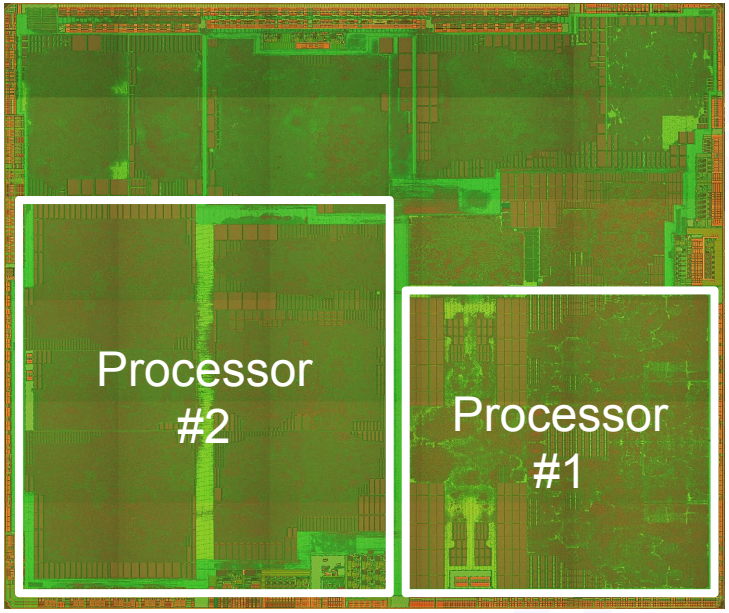


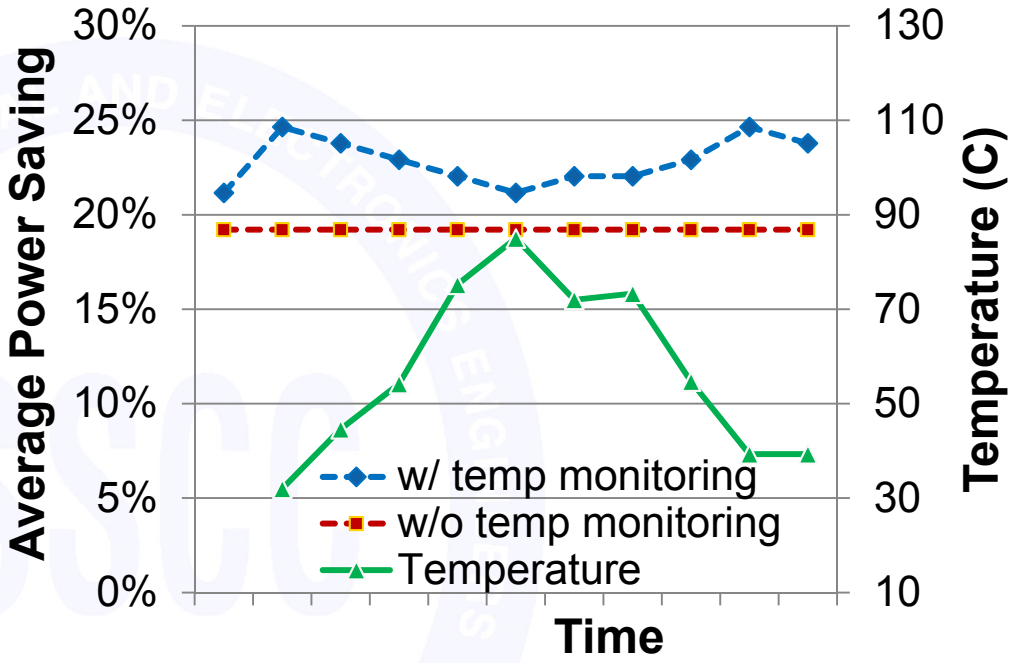




- CPU achieves +23% GHz or -30% average power savings
- GPU achieves +13% GHz and -20% average power savings

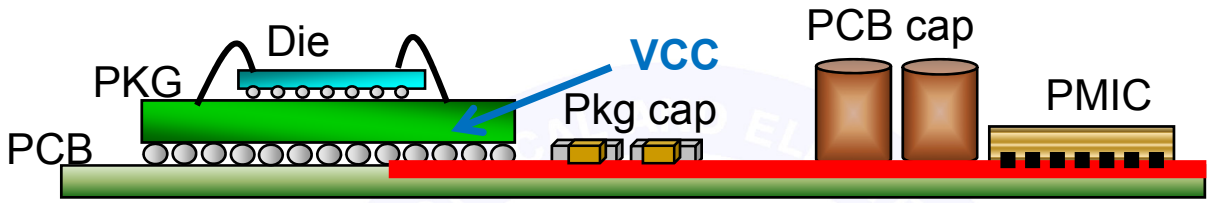
Temperature monitoring



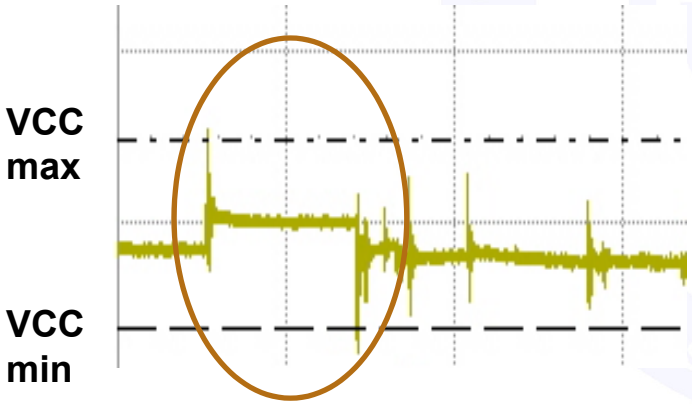


- Multiple temperature sensors distributed around the processors
- Interrupts from the PTP controller detect voltage adjustment needed
  - + Voltage to ensure MHz
  - - Voltage to save power
- Temperature monitoring achieves 2~5% additional power savings compared with no monitoring case

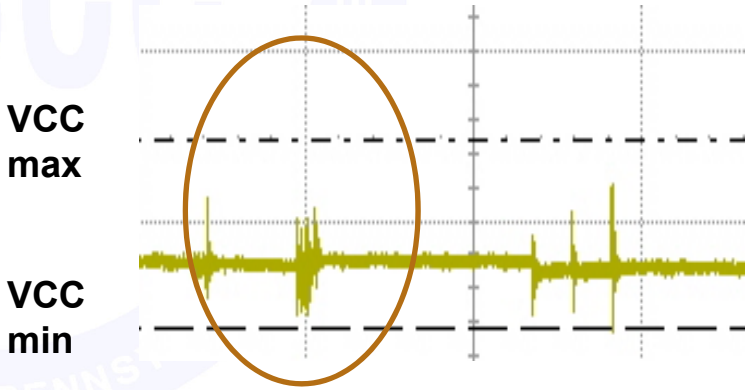
Remote sensing



- Continuously changing workload causes voltage droop on PDN and reduces voltage margin available
- A feedback line using PCB trace to a location closest to SoC pins provides actual SoC voltage to PMIC to compensate for the losses



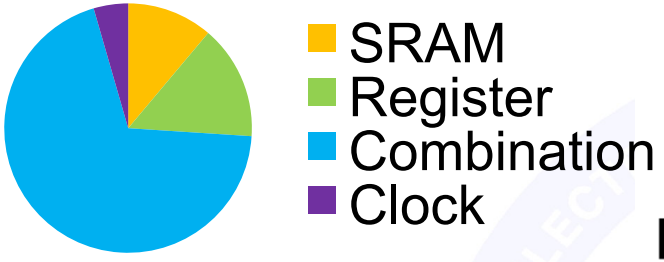
Remote sensing off



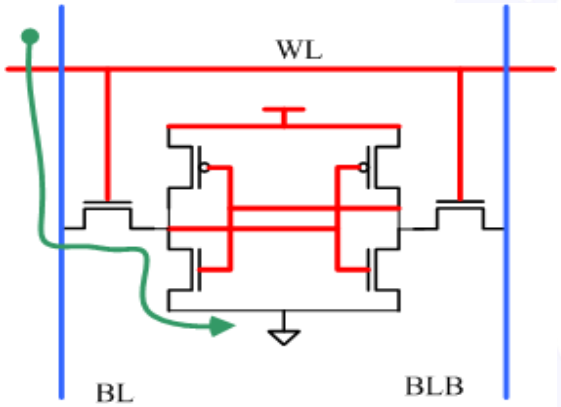
Remote sensing on

- 18mV(DC)/11mV(AC) VCC compensation with remote sensing

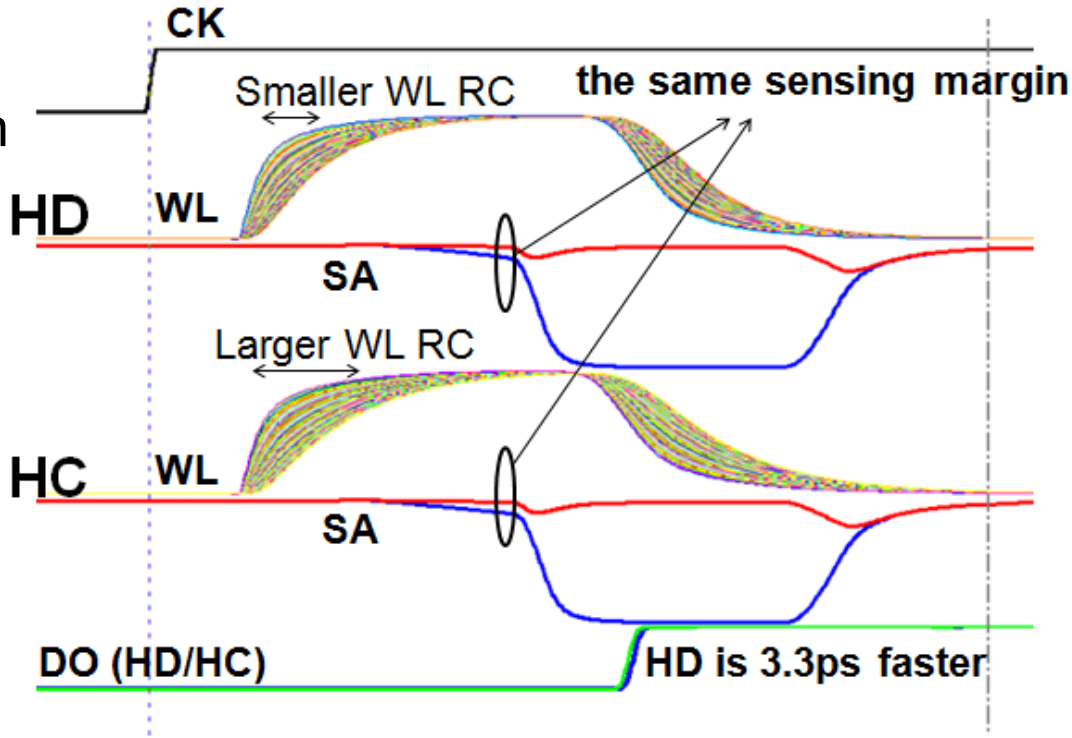
Low power SRAM (1/2)



Leakage power of CA15 on 28nm



Boosted WL and cell Vdd

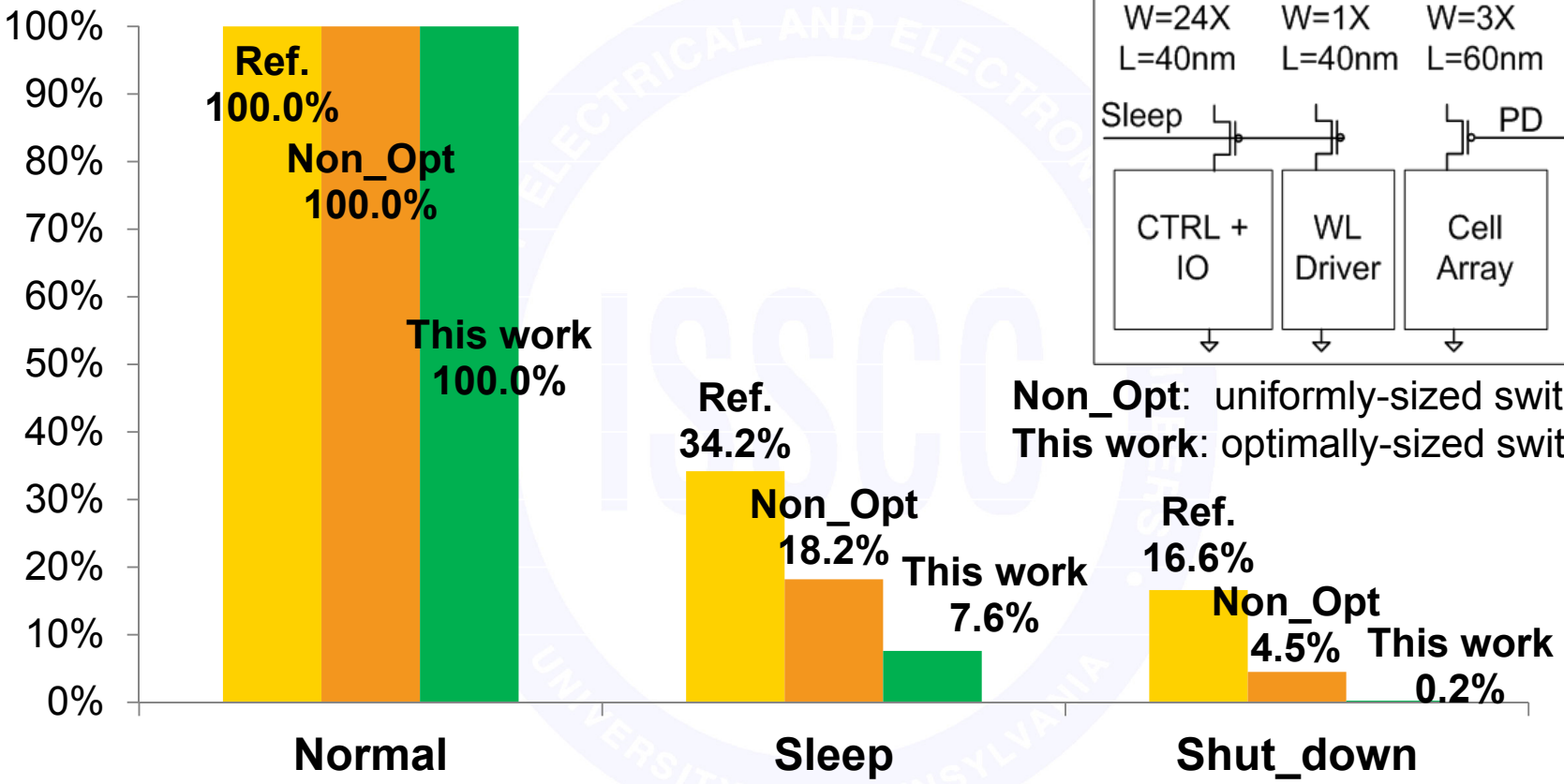


HD vs HC bit-cell

- HD (high-density) 6T bit-cell with boosted Vgs and pulldown devices for best power and performance balance
- 30% HD bit-cell leakage reduction (vs HC)

Low power SRAM (2/2)

Leakage reduction

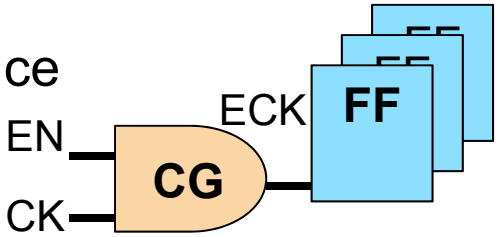


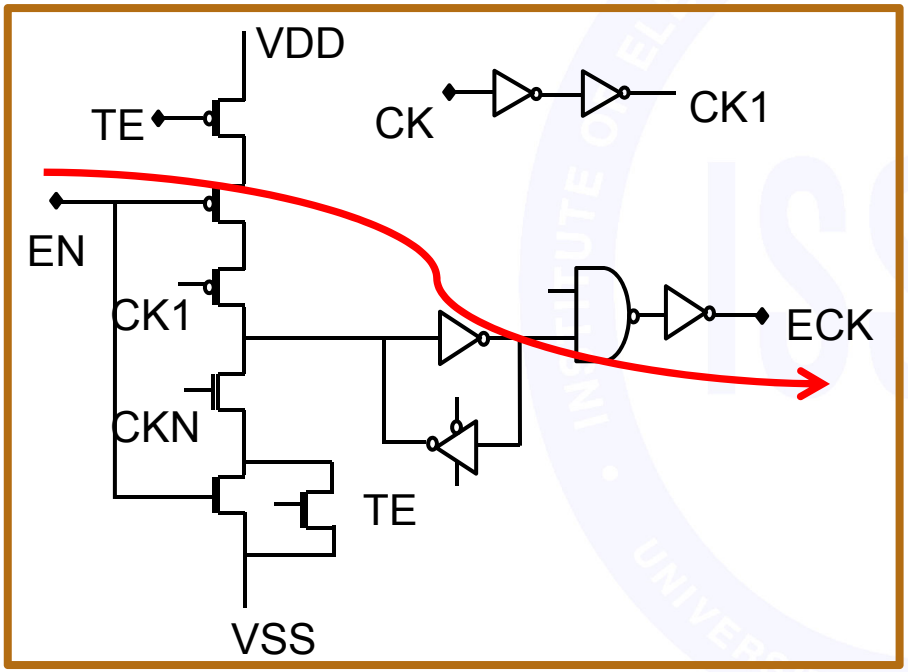
Non_Opt: uniformly-sized switches
This work: optimally-sized switches

- Optimally-sized power gating switches to balance speed, IR-drop and shutdown leakage
- Sleep : 13X ↓ , Shut_down : 425X ↓ (from Normal)

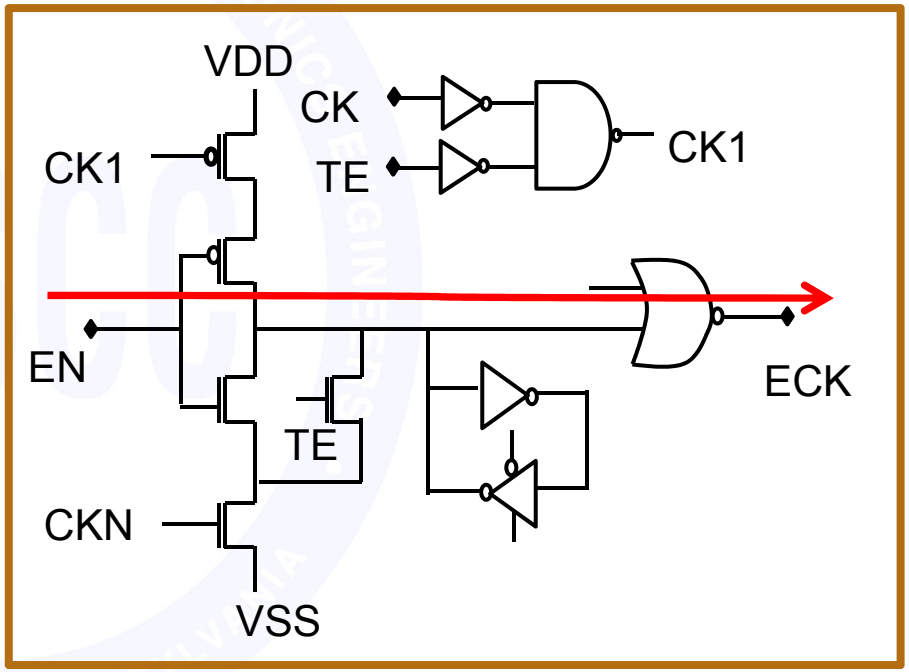
High speed CG cell (1/2)

- CG (clock-gating) is a common technique to reduce high CPU power by paying timing penalty





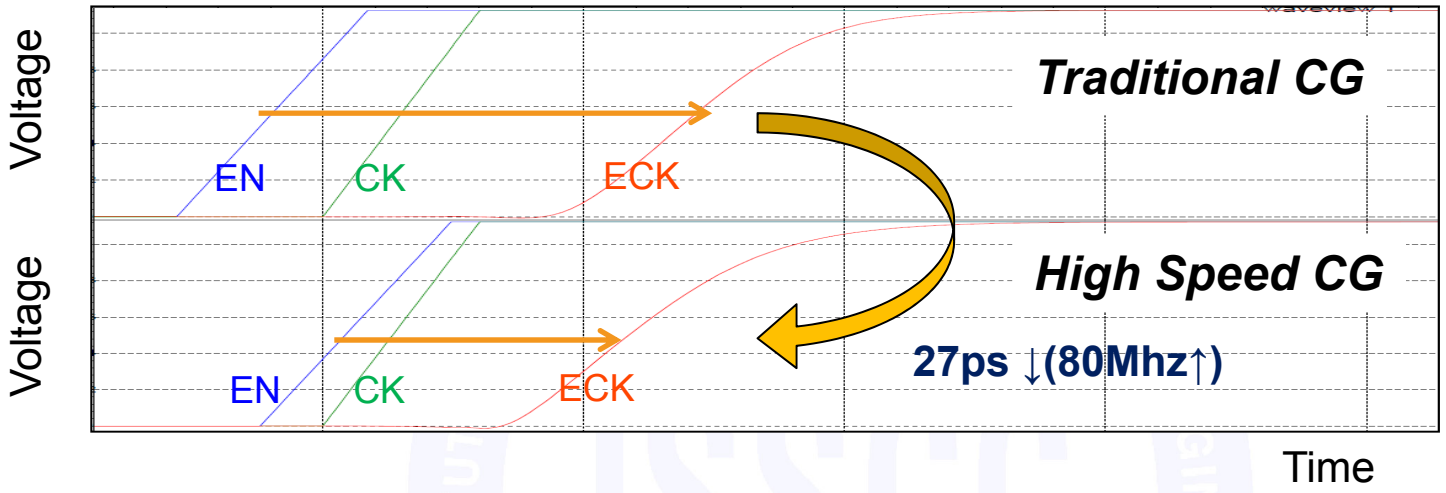
Traditional CG

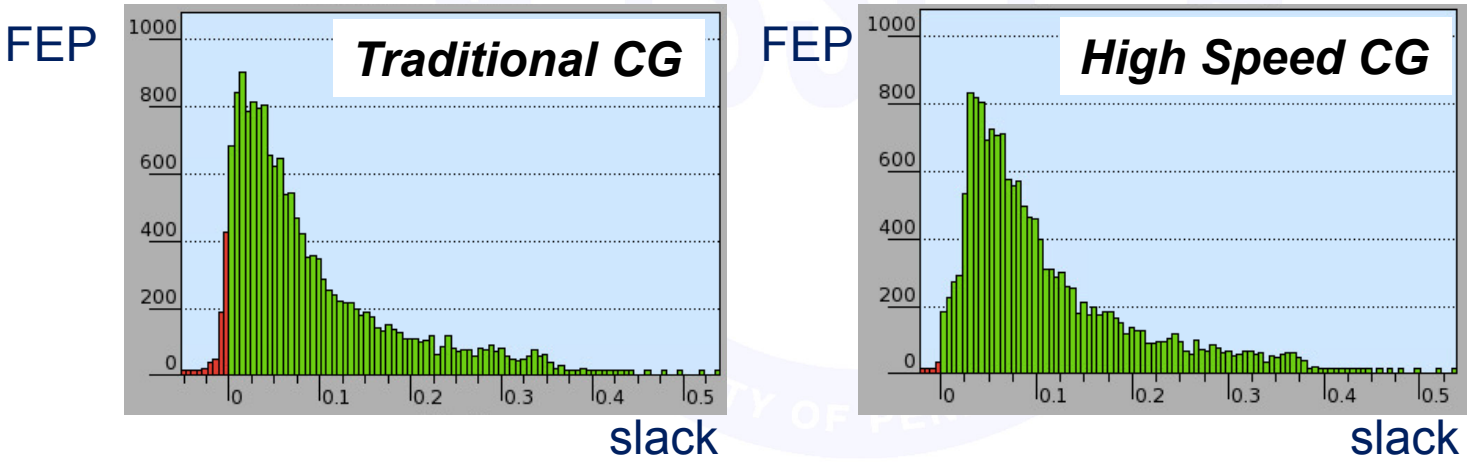


High Speed CG

- EN->ECK critical path optimized from 4 gates to 2 gates

High speed CG cell (2/2)





- Cortex-A15 STA with HS CG : **WNS: 27ps↓ TNS: 9X↓ FEP: 10X↓**

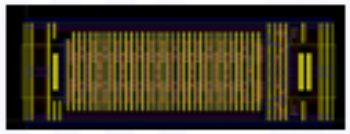
External shutdown/QH PS

Leakage at flight mode
(normalized)

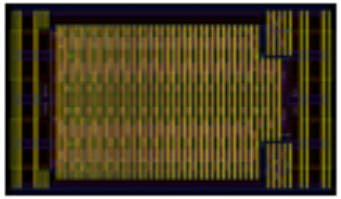
	Without external shutdown	With external shutdown
CPU	26.4%	6.6%
GPU	24.0%	12.7%
Others	49.6%	49.6%
SoC total	100.0%	68.9%

Quad height (QH) vs Double height (DH) power switches

	Area [um2]	Ron [ohm]	Ion [uA]	Ion/Area [uA/um2]
QH/DH	2.14X	32.9%	3X	1.41X



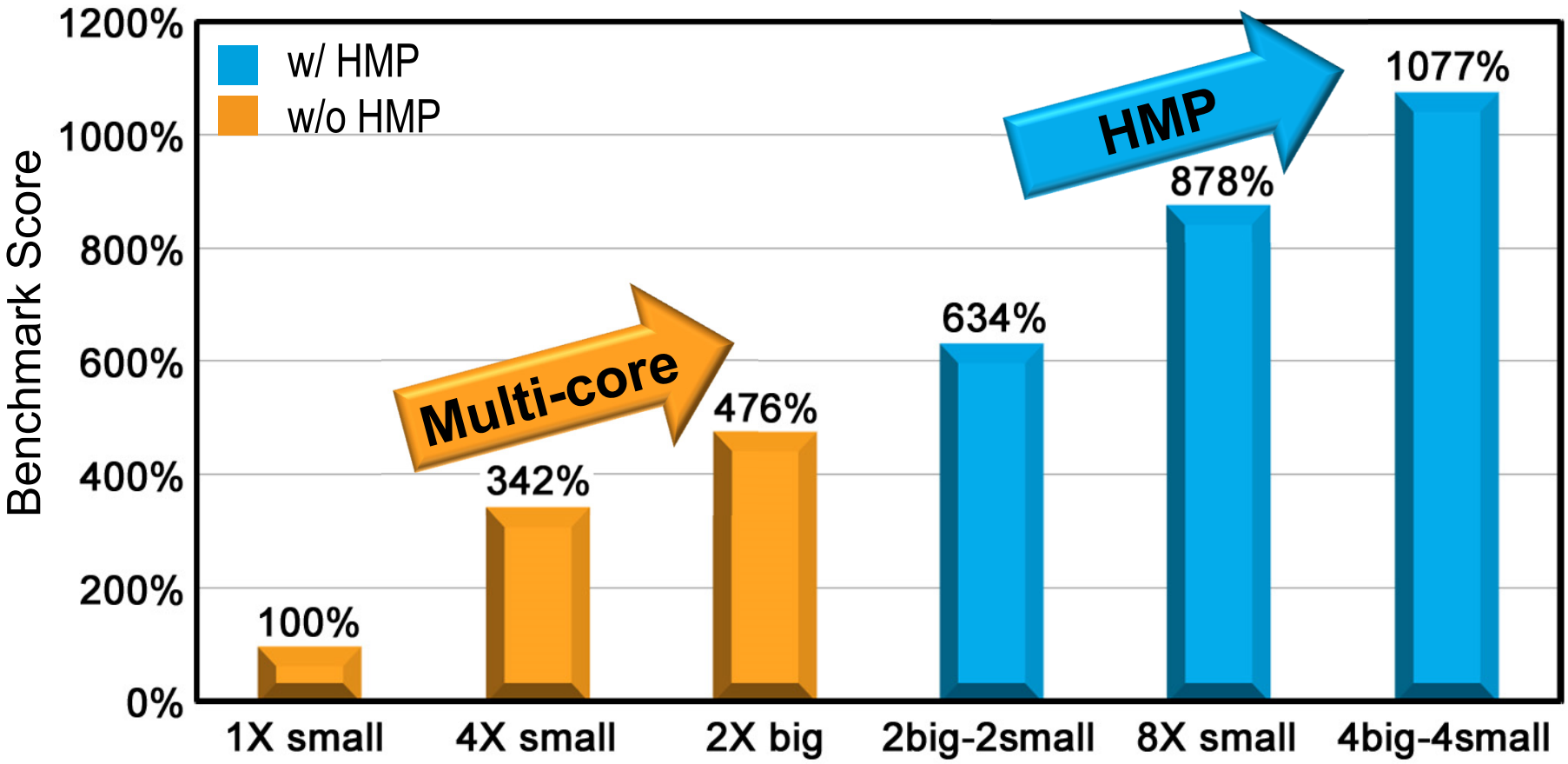
Double-height (DH) power switch



Quad-height (QH) power switch

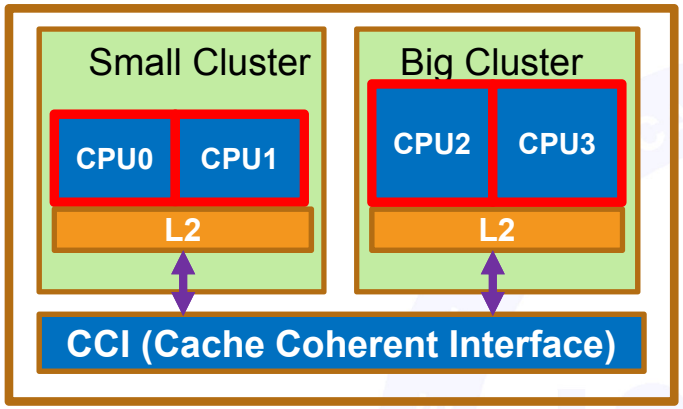
- PMIC shuts down power switches and AO in CPU and GPU, 31% leakage reduction at flight mode
- QH PS improves 41% uA/um2 and saves CPU switch area by 2X while meeting the same IR drop spec

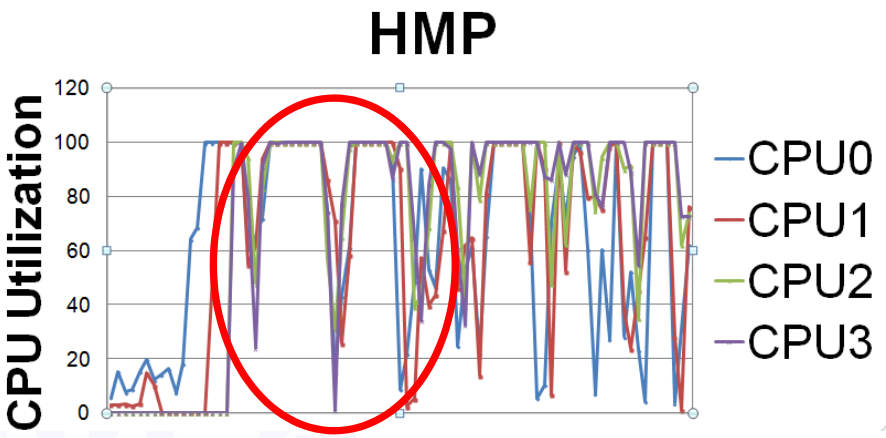
Multi-core Architecture and System Optimization

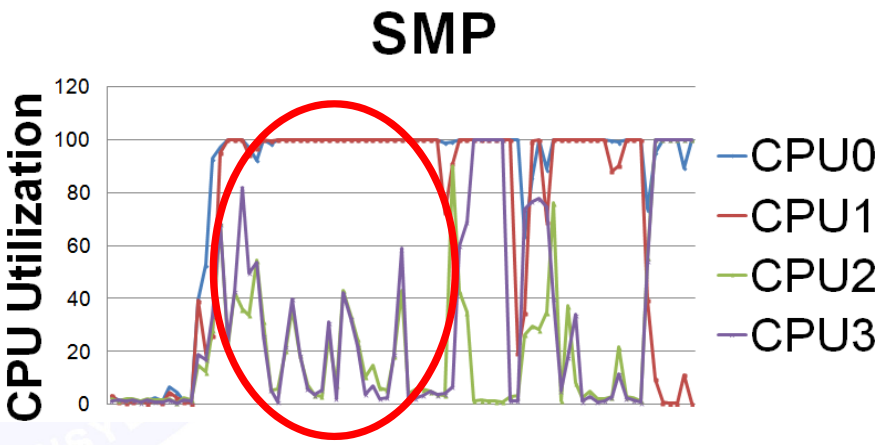


- HMP optimizes performance (big CPU) and energy efficiency (small CPU)
- A good HMP model is needed for maximize performance beyond multi-core

HMP (Heterogeneous Multi-Processing)



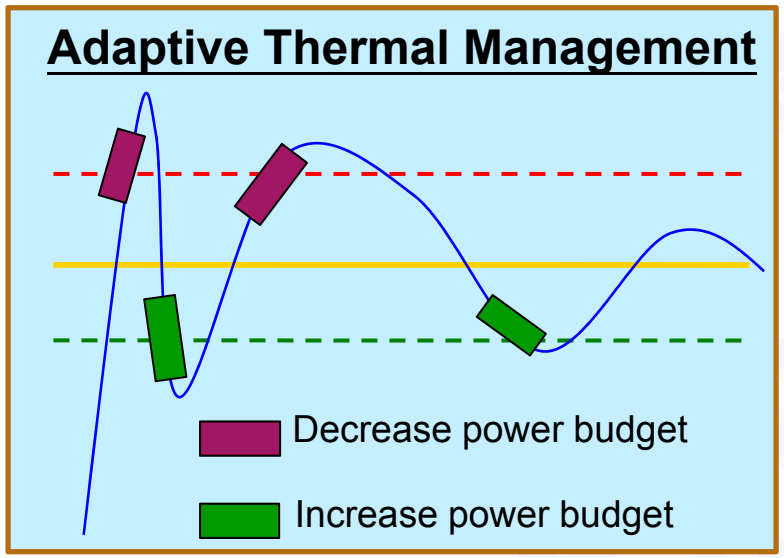




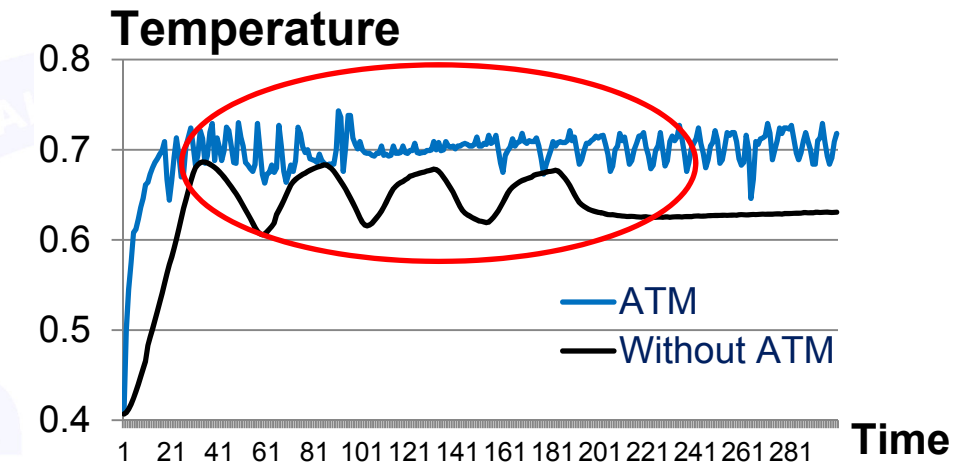
	Cluster Migration	CPU Migration	SMP	HMP
Switching granularity	Cluster	CPU	Unrestricted	Unrestricted
Flexibility	Low	Medium	High	High
Max perf.	2xBig	2xBig	2xBig+ 2xSmall	2xBig+ 2xSmall
Power saving	Low	Medium	Low	High

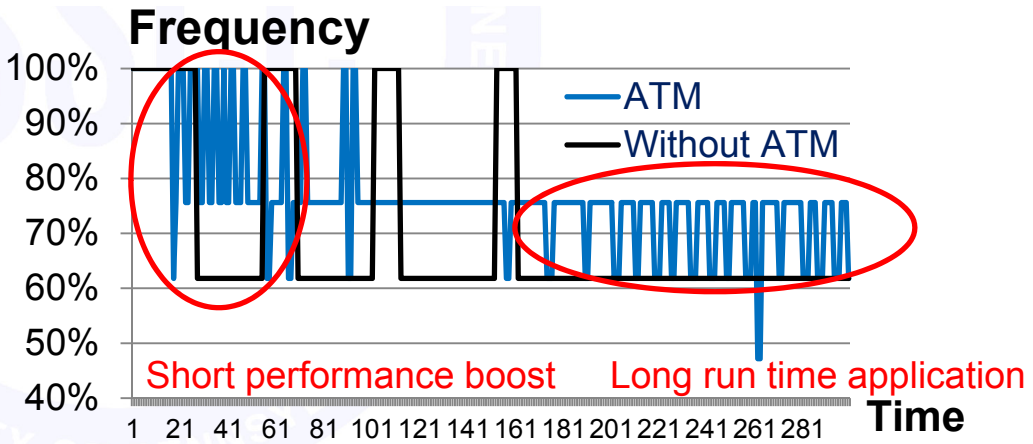
- HMP increases performance by 33~51% during “heavy workload” and increases energy efficiency by 2~5X during “light workload”

ATM (Adaptive Thermal Management)



▪ATM takes the input from thermal sensors and outputs the decreased/ increased power budget to IPM based on the slope of rise/fall temperature





- ATM offers 10% performance uplift over traditional fixed table thermal throttling

Summary

- **Circuit/Architecture/System** solutions needed to close the increasing energy gap

- **PTP Tech** +23%Ghz or -30% power
  Remote Sensing -18mV(DC)/11mV(AC) variation
  LP SRAM 13X(sleep)/425X(shut_down) leakage
  HS CG WNS 27ps,TNS 9X,FEP 10X
  External shutdown -31% leakage
  Quad-height PS 2X PS area
  HMP +33~51% performance
  ATM +10% performance
  Co-optimization of all improves performance, power and thermal to provide the best user experience

A 1.22TOPS and 1.52mW/MHz Augmented Reality Multi-core Processor with Neural Network NoC for HMD Applications

Gyeonghoon Kim, Y. Kim, K. Lee, S. Park, I. Hong, K. Bong, D. Shin, S. Choi, J. Oh and Hoi-Jun Yoo

Dept. of EE, KAIST

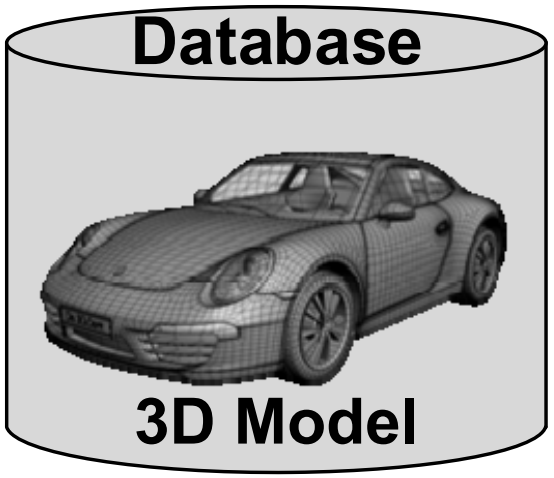
Outline

- Markerless Augmented Reality (AR) in HMD
- Block Diagram of the Proposed Processor
- Key Features
  - Task-level Pipelined Many-SIMD Architecture
  - Neural Network Scheduled 2D-mesh NoC
  - Mixed-mode SVM-based Dynamic Resource Management
- Implementation Results
- Conclusion

gmented Reality (AR) ?

Recognize

Visualize

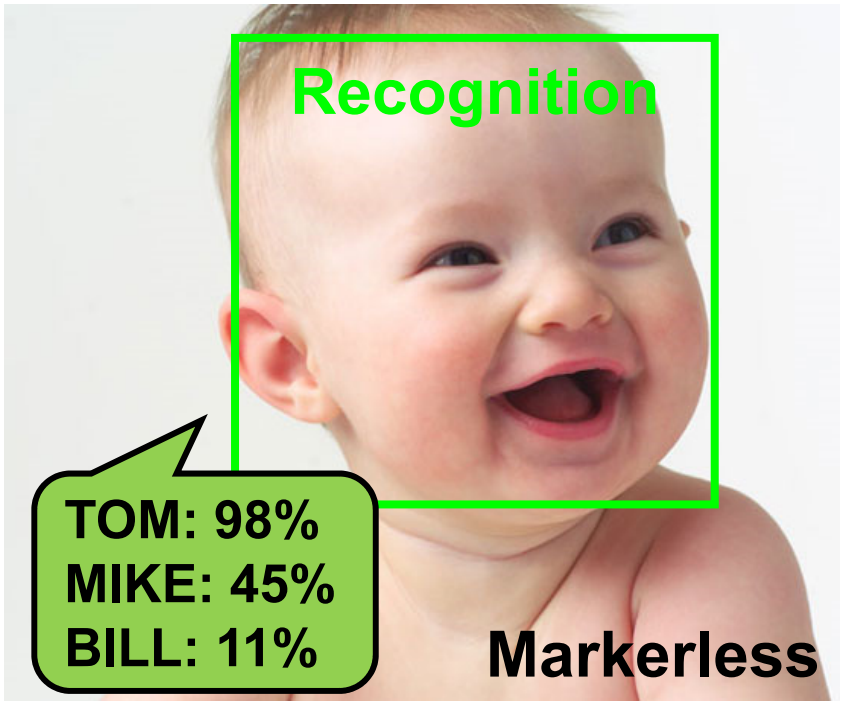


Real-world Environment

Augmented Reality (AR)

Virtual Reality

1) Requirements for AR in HMD: High-Throughput



- 2D-barcode / 2D-marker
- Markers on Everything?

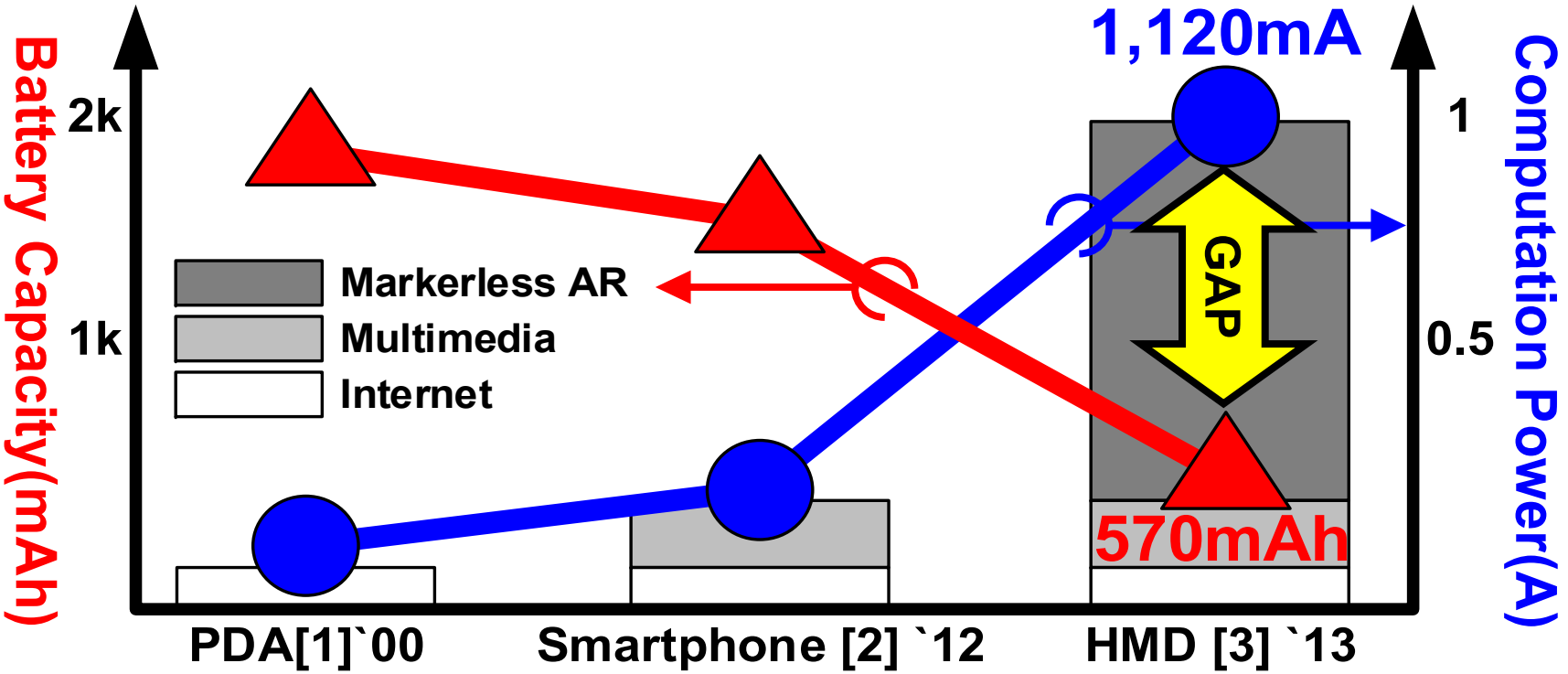
➔ Impossible Solution in Real World

- Natural Feature Extraction
- ~10x Higher Computation

➔ Difficulties in Real-time (30fps) Operation

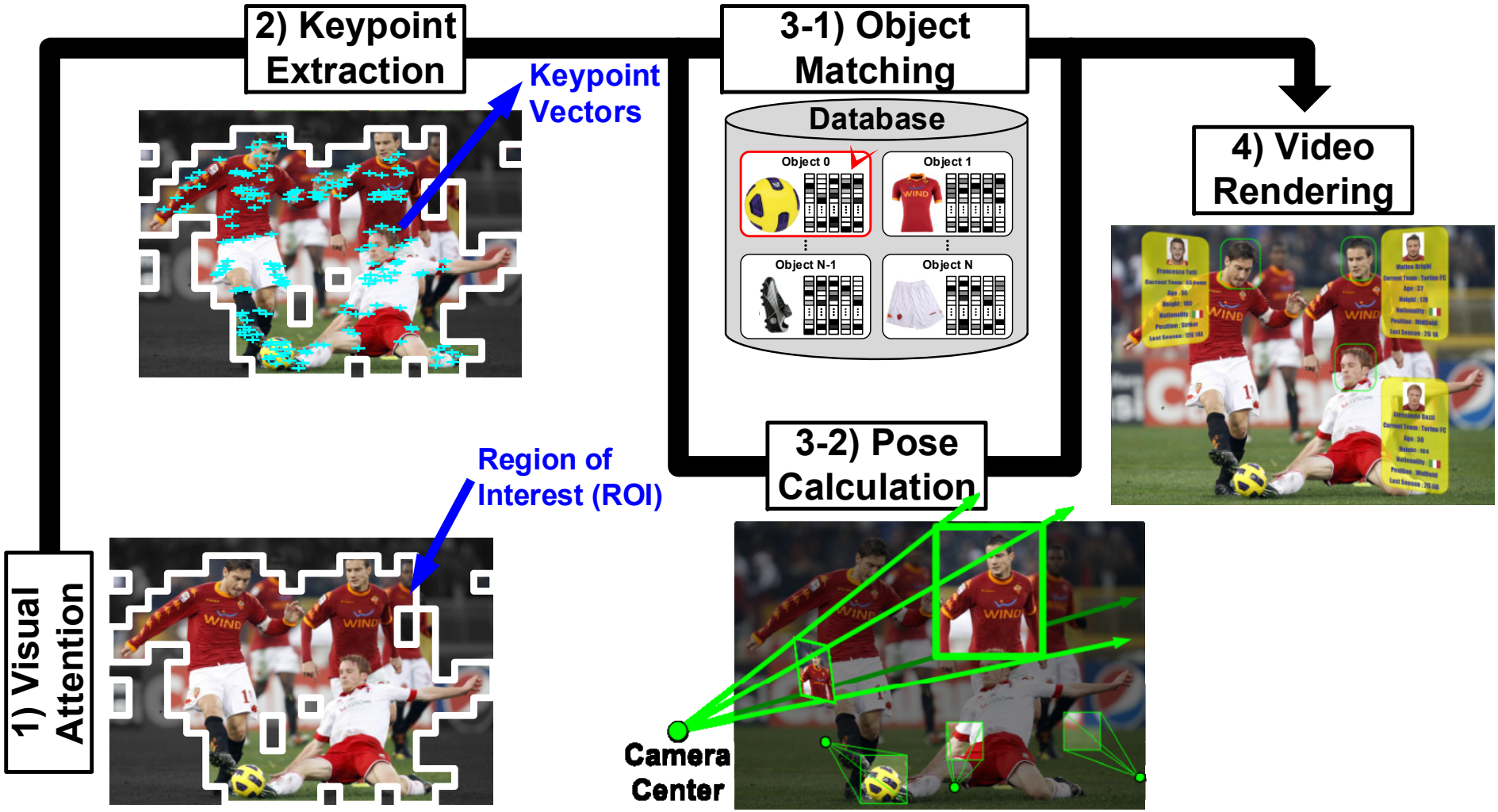
2) Requirements for AR in HMD: Low-Energy

- Hands-free Head Mounted Display (HMD)
  - Smaller Battery Capability
  - Complicated Markerless AR Functionalities

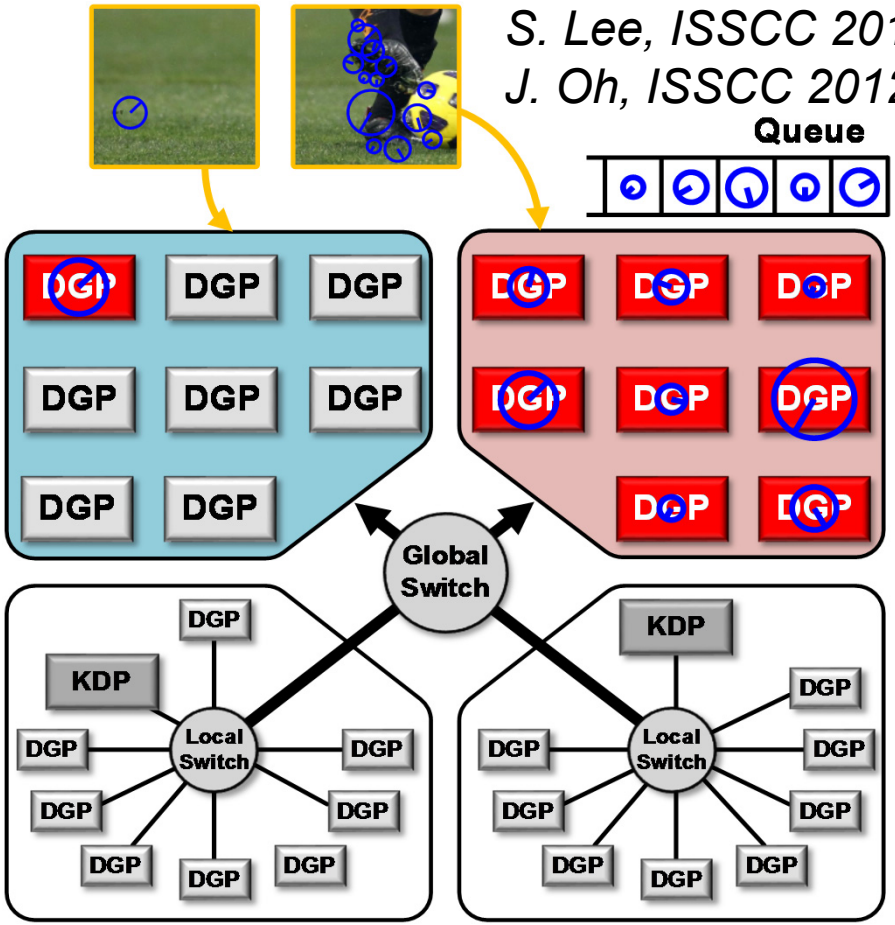


[1] HP iPAQ [2] Apple iPhone5 [3] Google Project Glass

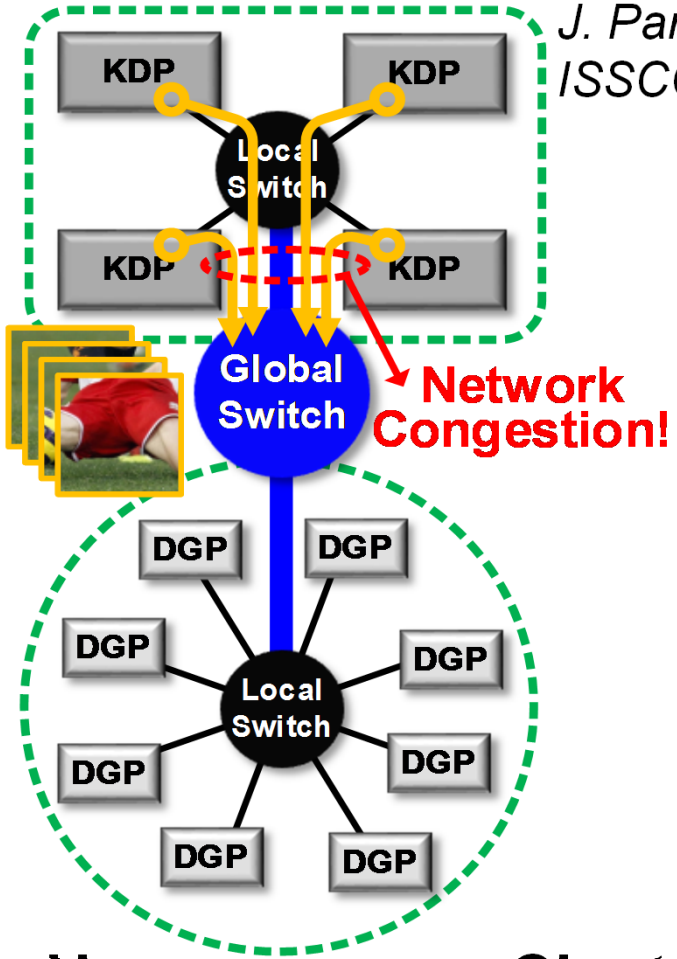
Augmented Reality Algorithm Flow



Previous Vision Processor Architecture



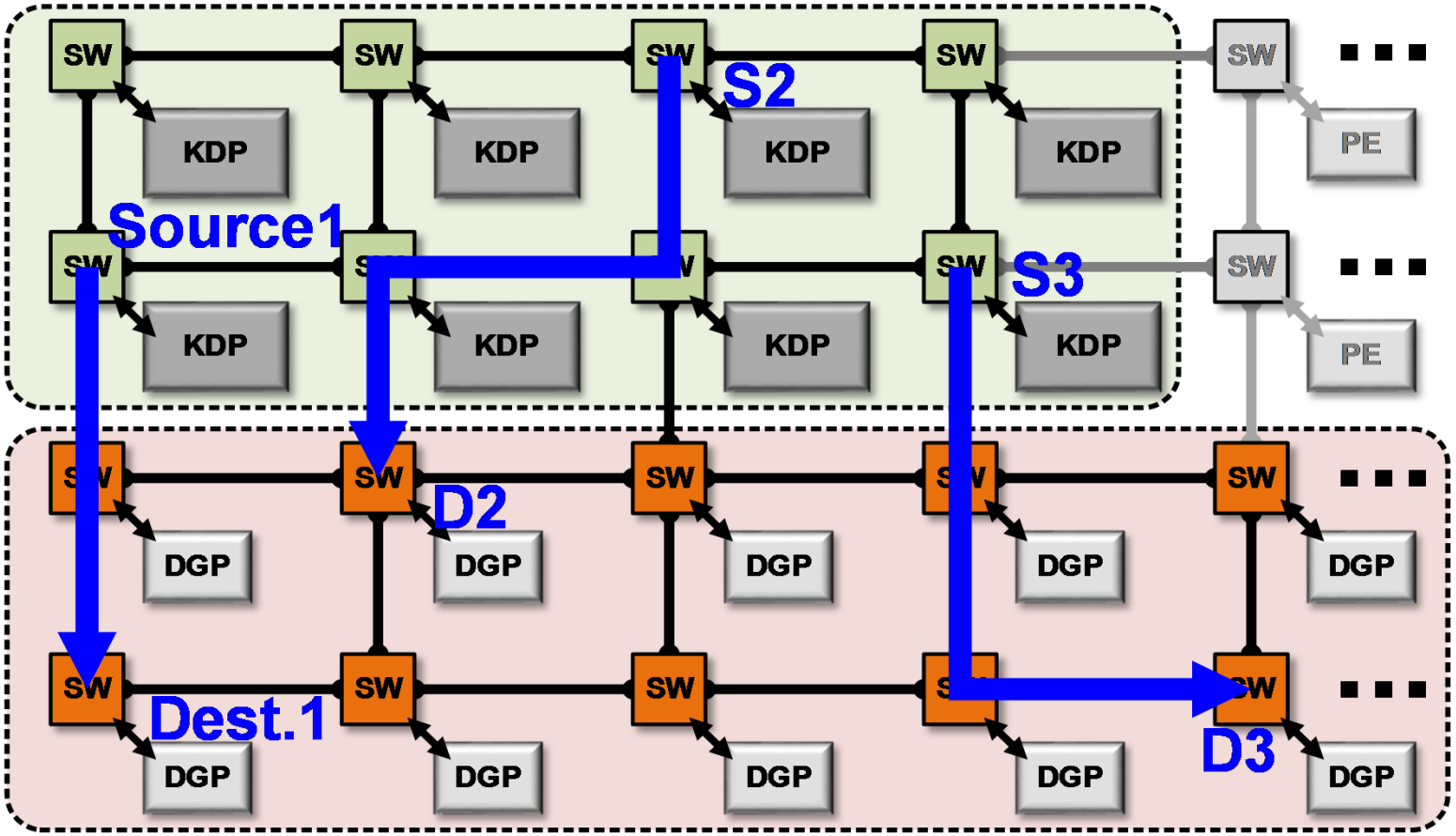
- Heterogeneous Cluster w/ Hier. Star NoC
- Uneven Workload



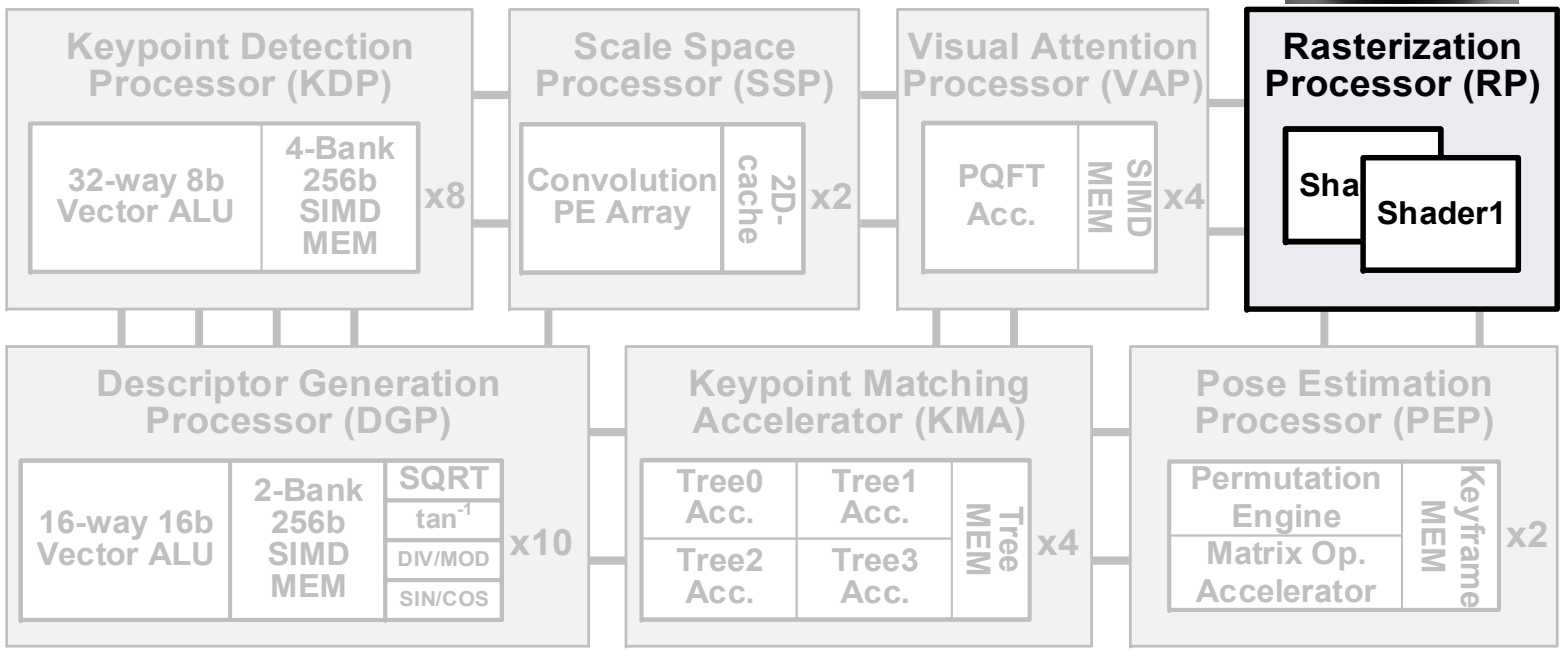
- Homogeneous Cluster w/ Hier. Star NoC
- Network Congestion

Homogeneous Cluster w/ 2D-mesh NoC

- **High Utilization** of Processors
- **Multi-source Multi-destination Data Transfer**



Overall Architecture of AR Processor



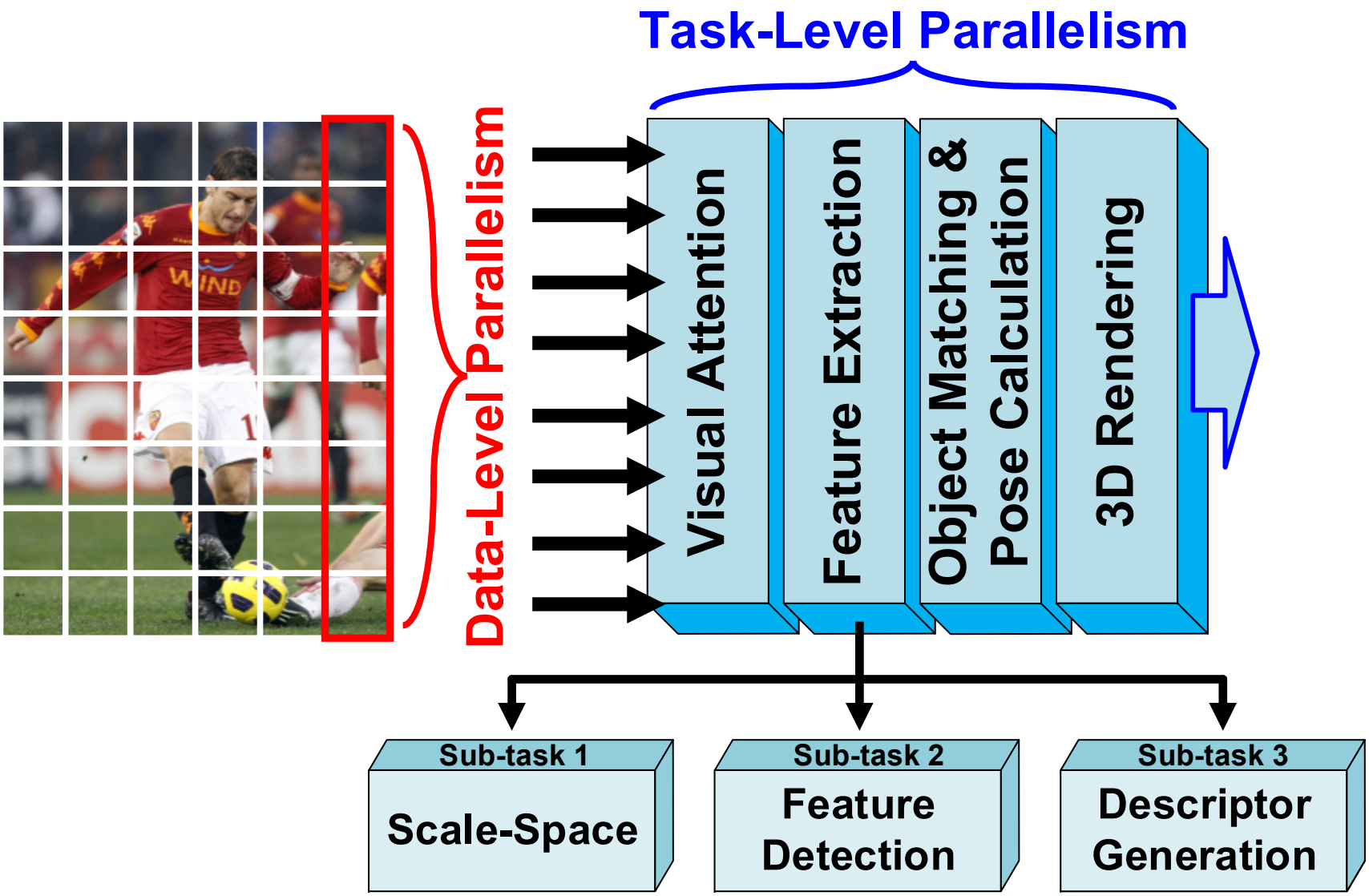
Key Features

1. For High-Throughput
   ➔Task-level Pipelined Many-SIMD Architecture

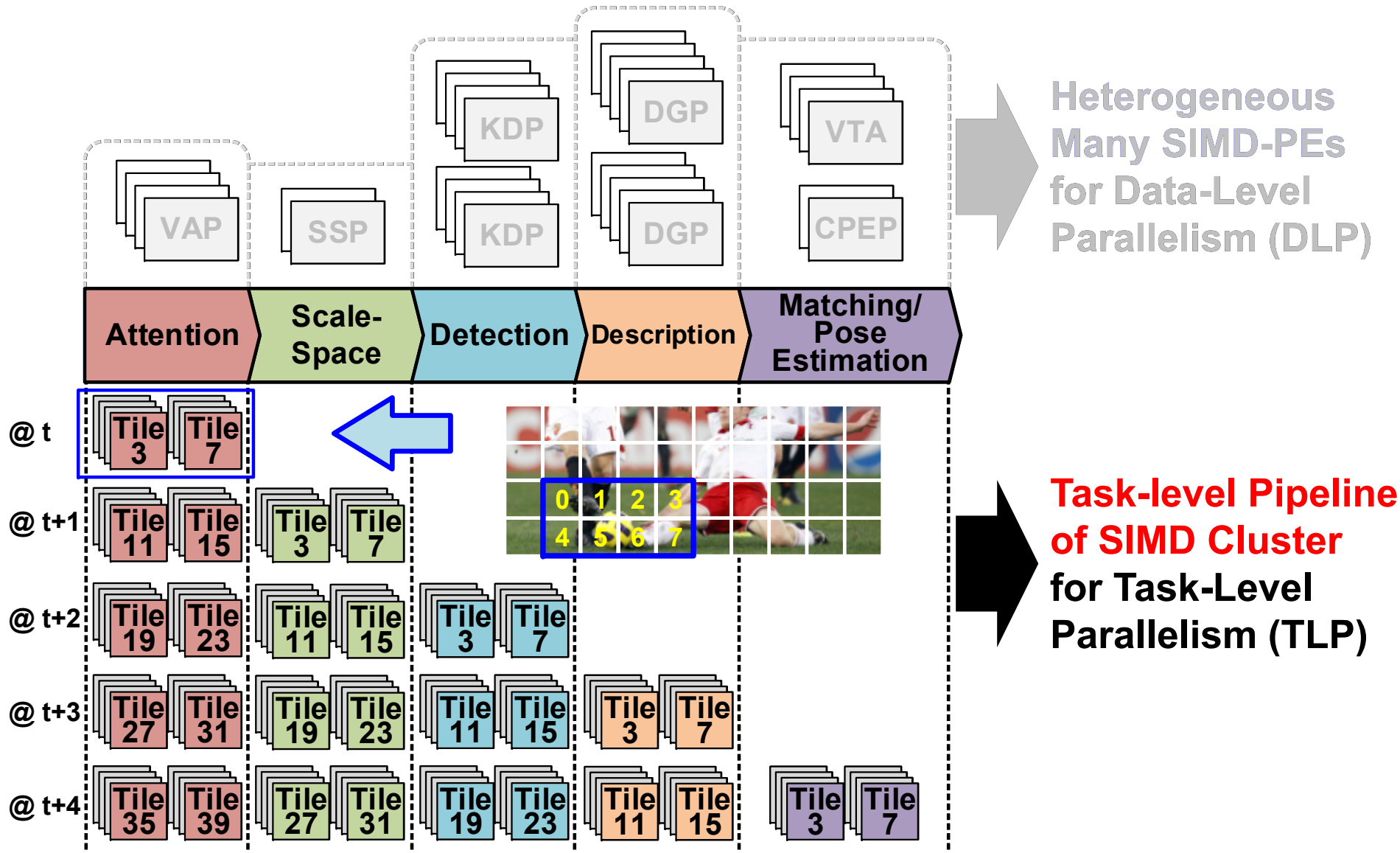
2. For Low Network Congestion
   ➔ Neural Network Scheduler for 2D Mesh NoC

3. For Low Energy Consumption
   ➔ Mixed-Mode Support Vector Machine (SVM)-based Dynamic Resource Management (DRM)

Data/Task-Level Parallelism in AR

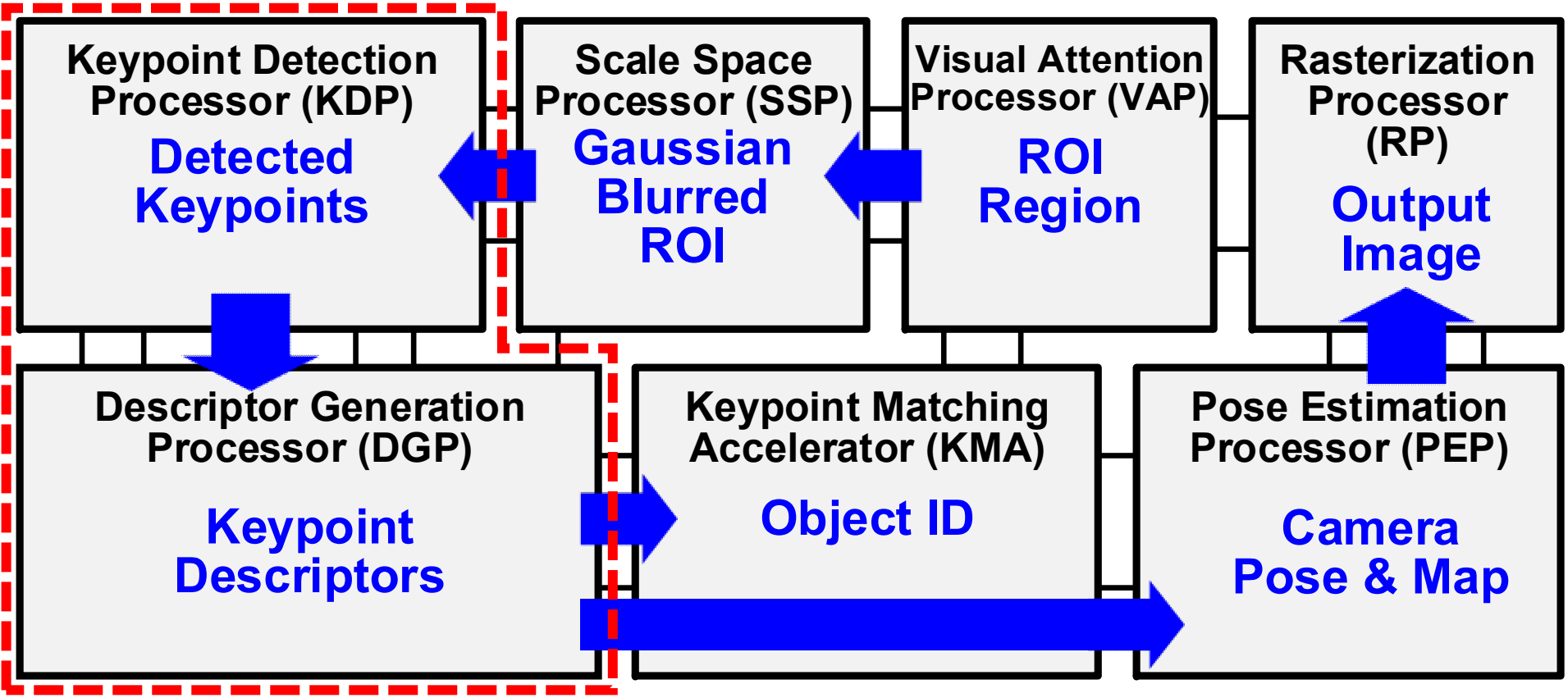


Task-level Pipelined Many-SIMD Processors



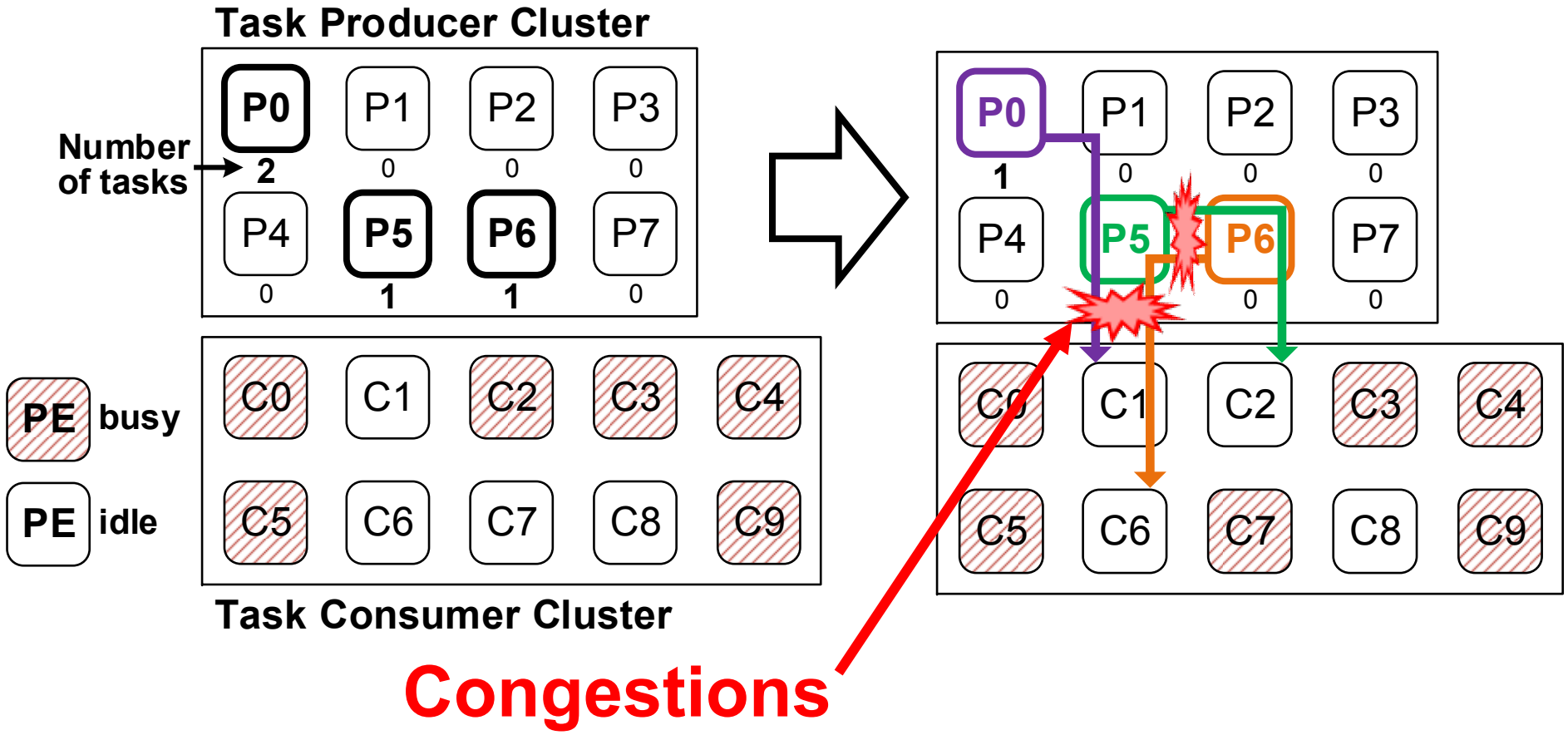
Problem in Task-level Pipeline Architecture

- **Simultaneous & Massive** Internal Data Transfer Between Clusters ➔ **Network Congestion!**



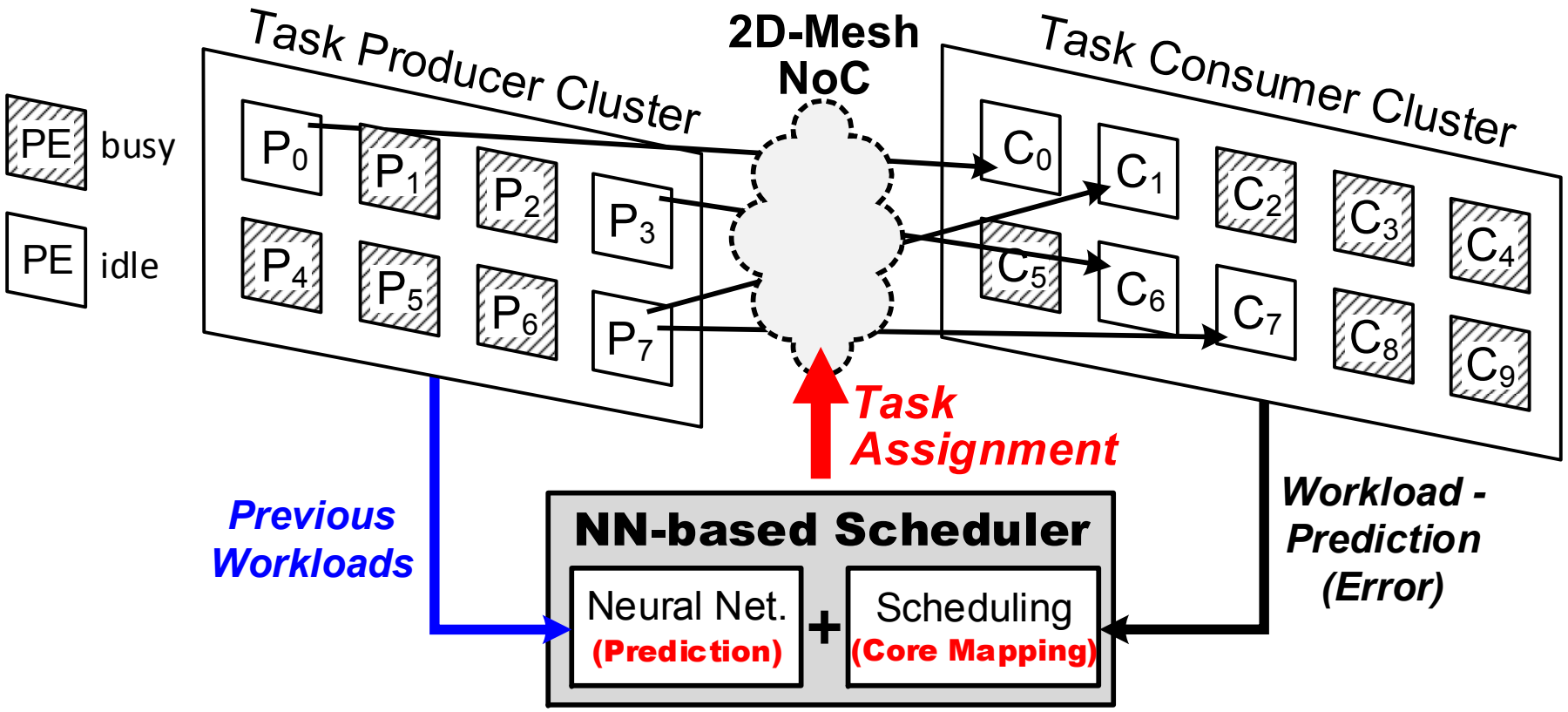
Network Congestion in 2D-mesh NoC

❑ Greedy Task Assignment w/o Workload Prediction

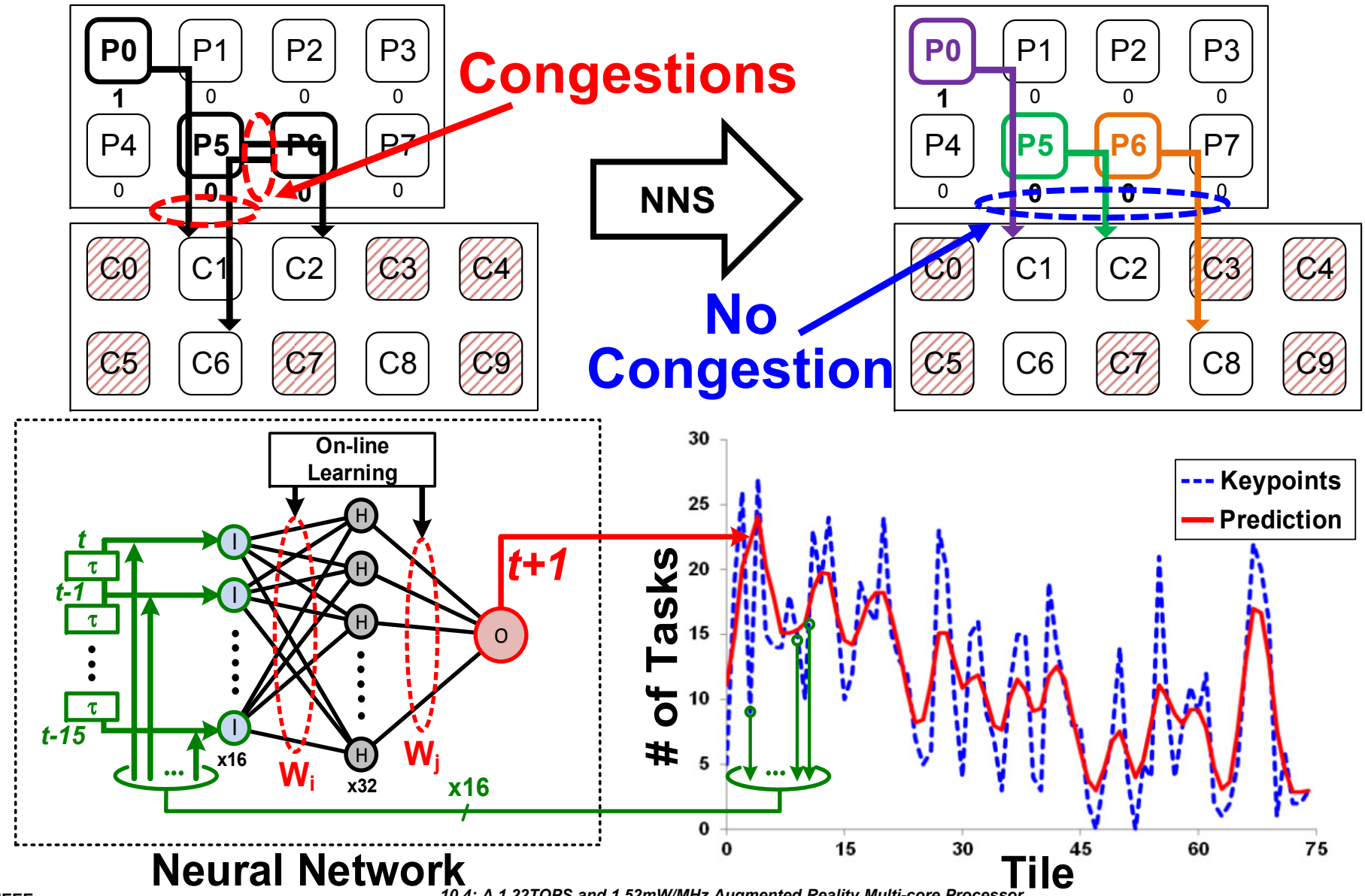


Neural Network Scheduler (NNS)

- Network Congestion Depends on Routing Path
- Prediction-based Task Assignment to Minimize Congestion ➔ 24.4% Shorter Latency

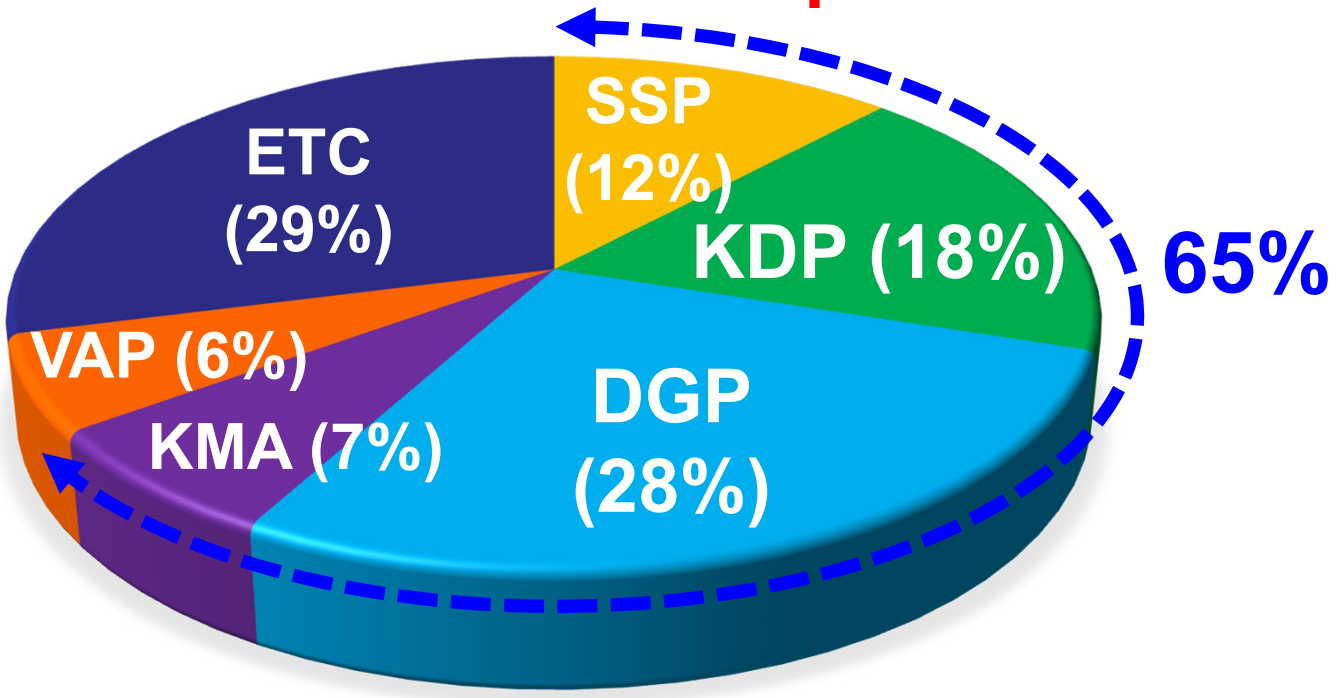


Neural Network Scheduler for mesh NoC



Overall Power Breakdown

- **Local Feature Extraction Processors Consume 65% of Total Power Consumption**

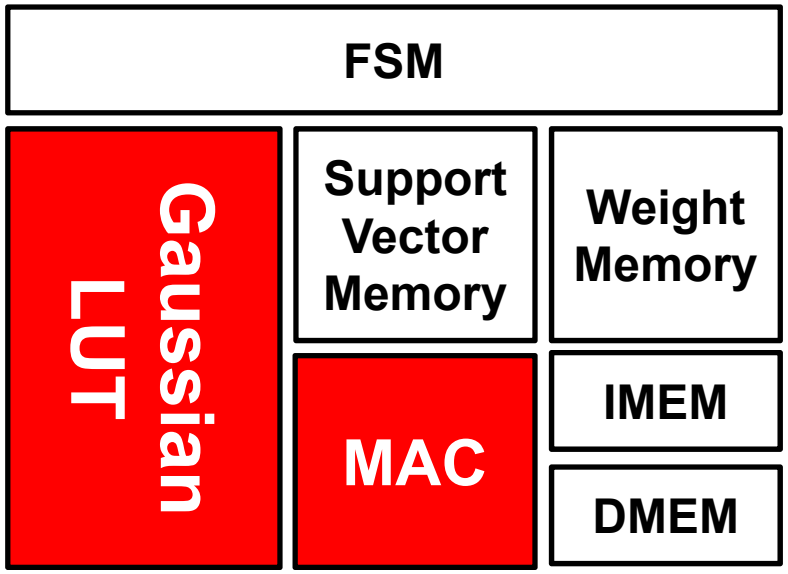


- **Workload Parameter 1: Number of Selected ROIs**
- **Workload Parameter 2: Number of Extracted Features**
- **Workload Parameter 3 & 4 : Thermal & Power Headroom**

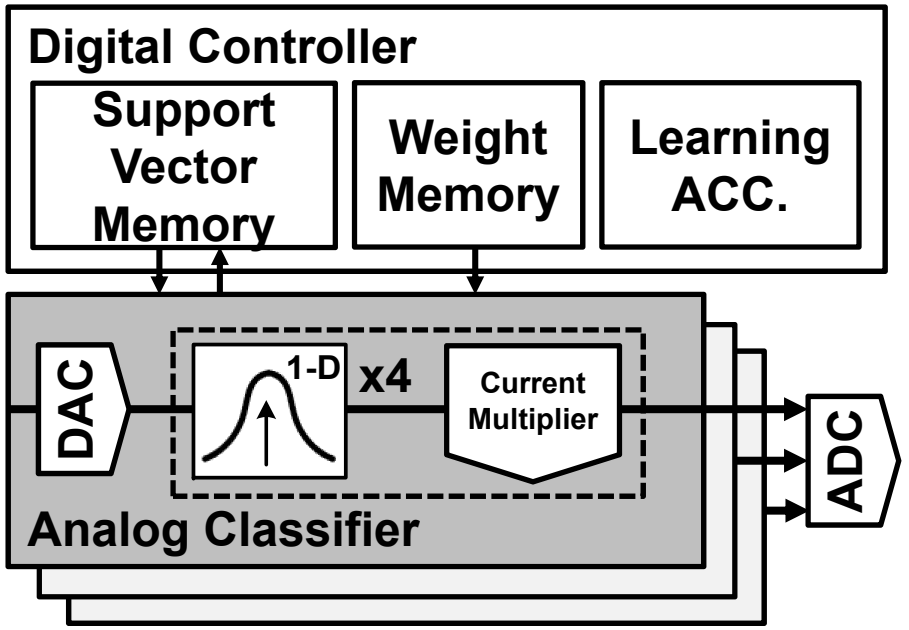
Mixed-mode SVM Implementation

- 2.7x Power Efficiency, 1.6x Area Efficiency

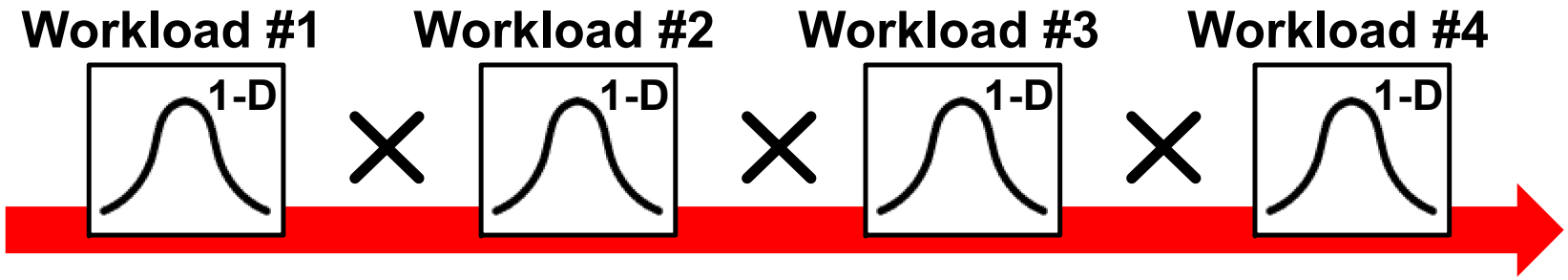
Digital SVM Classifier

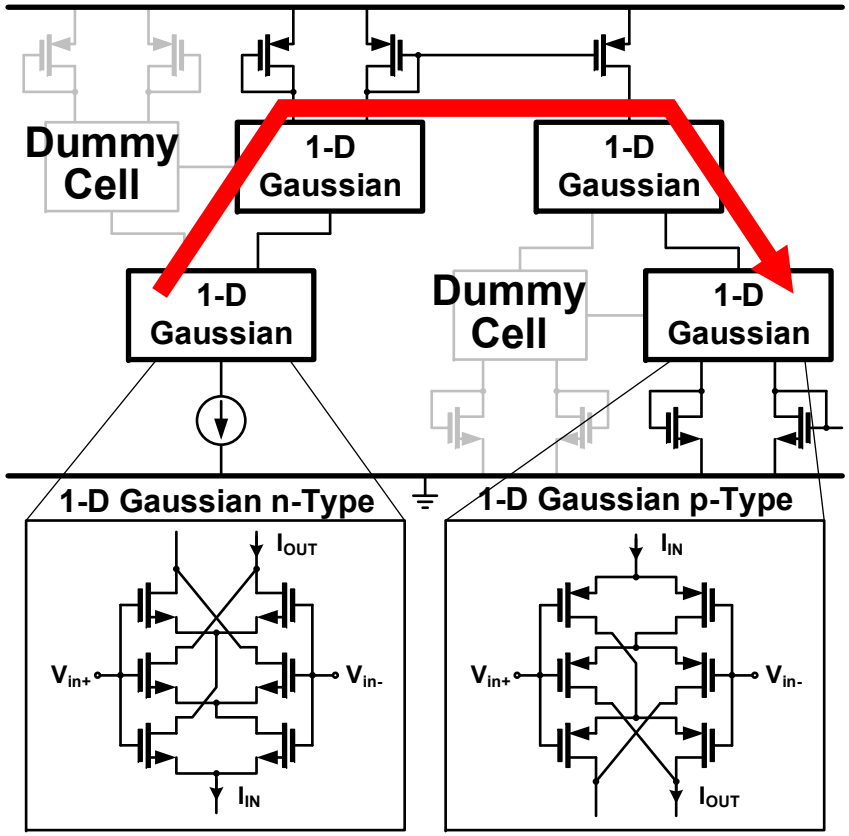


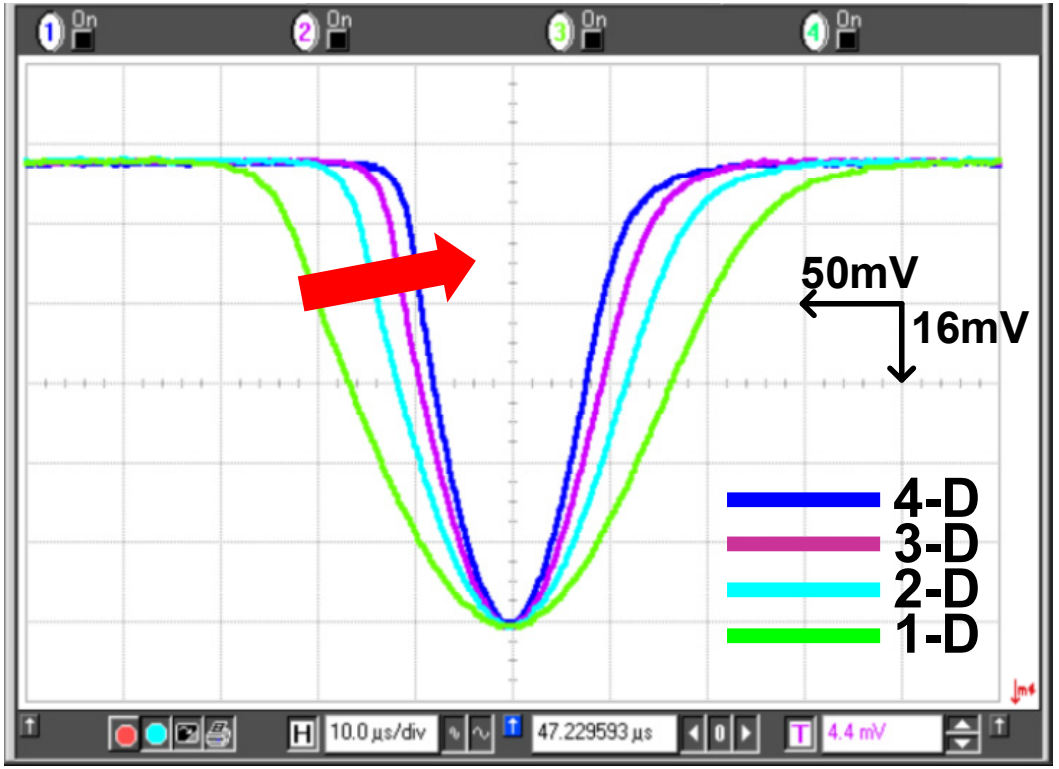
Mixed-mode SVM Classifier



4D-Gaussian for 4 Workload Parameters

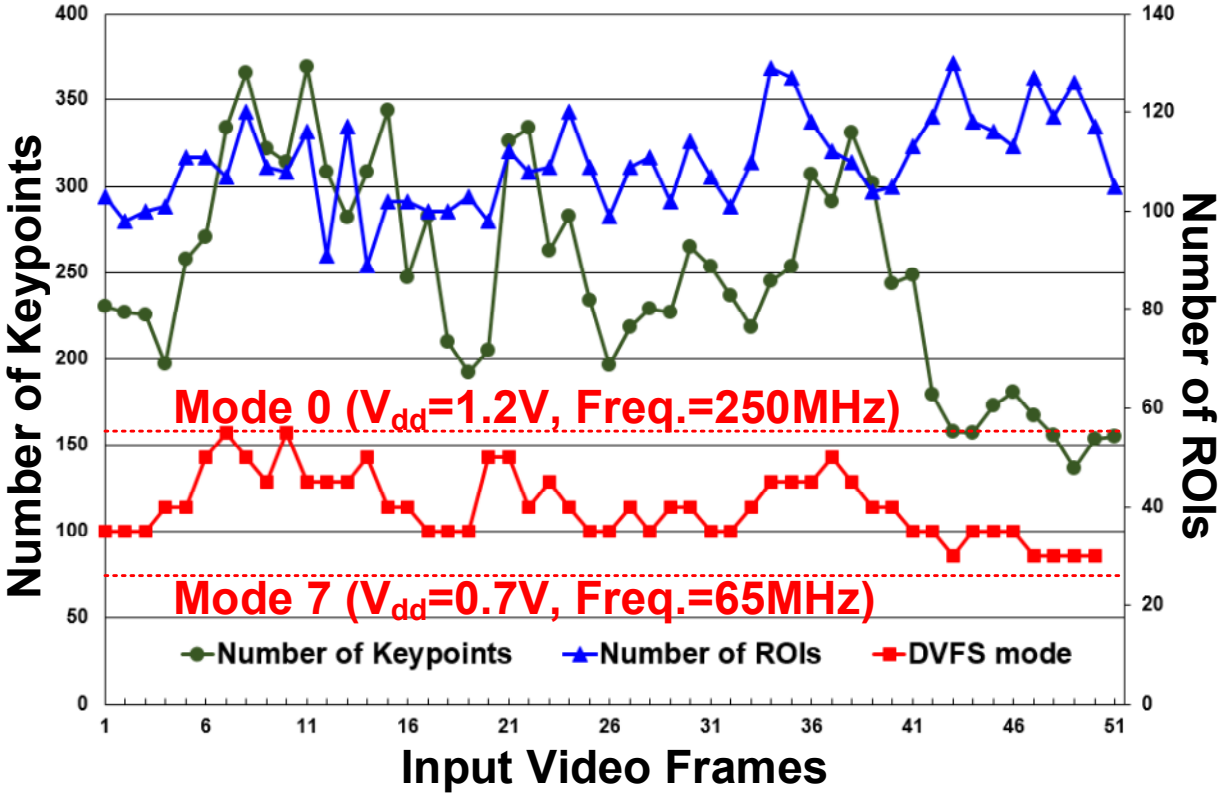


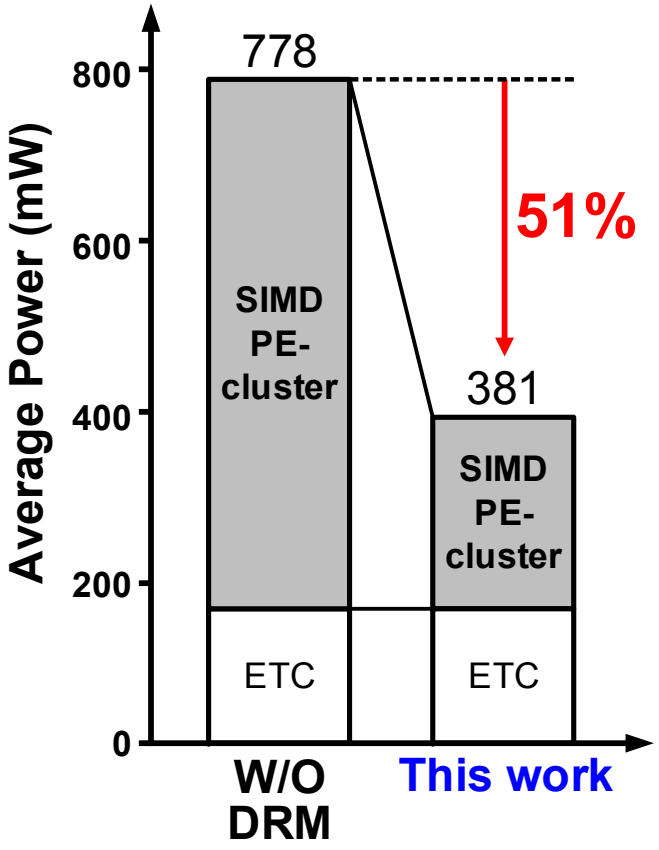




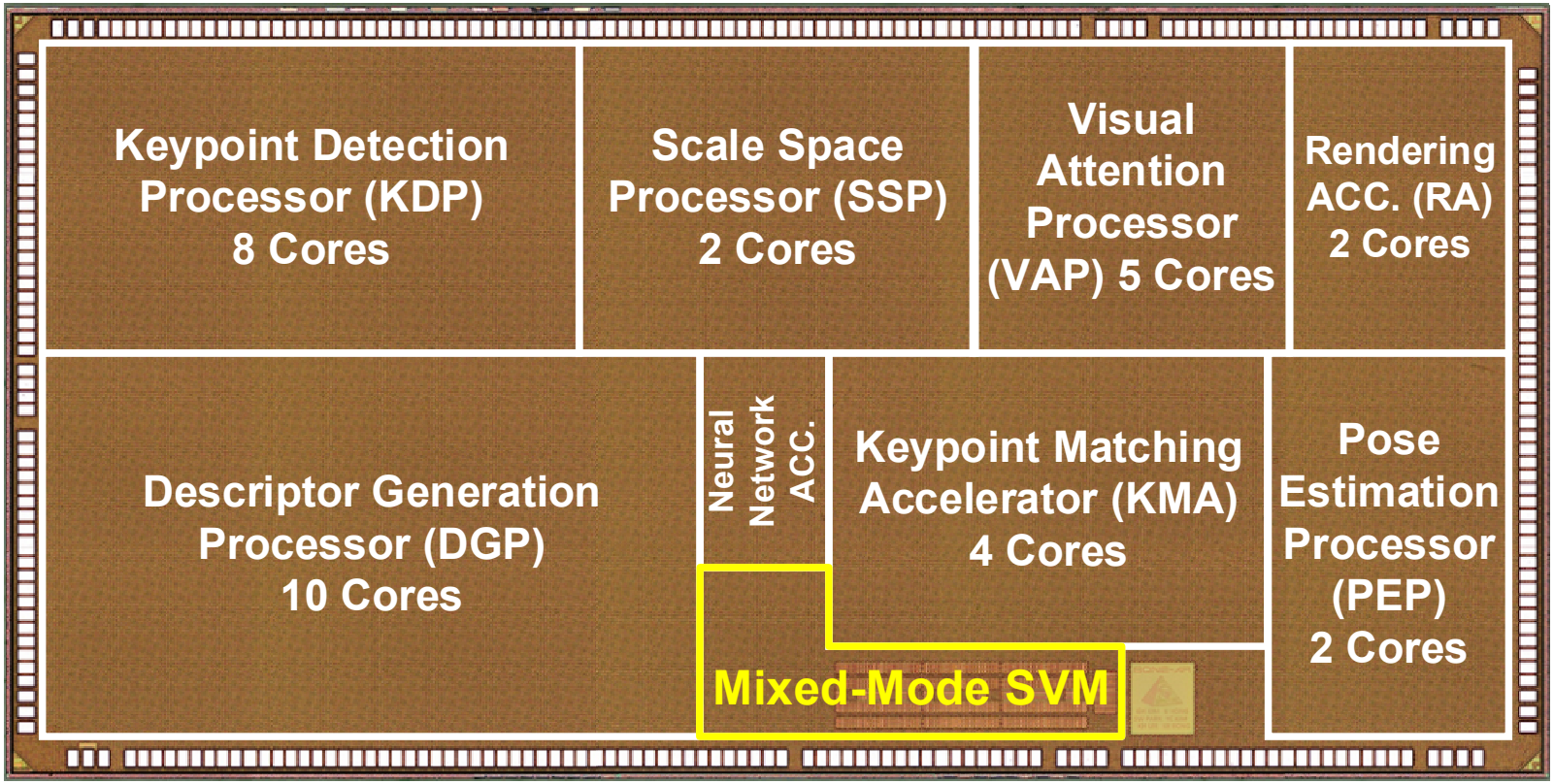
SVM-based Dynamic Resource Management

- Per-frame Power-mode Control
- 51% Reduction in Power Consumption



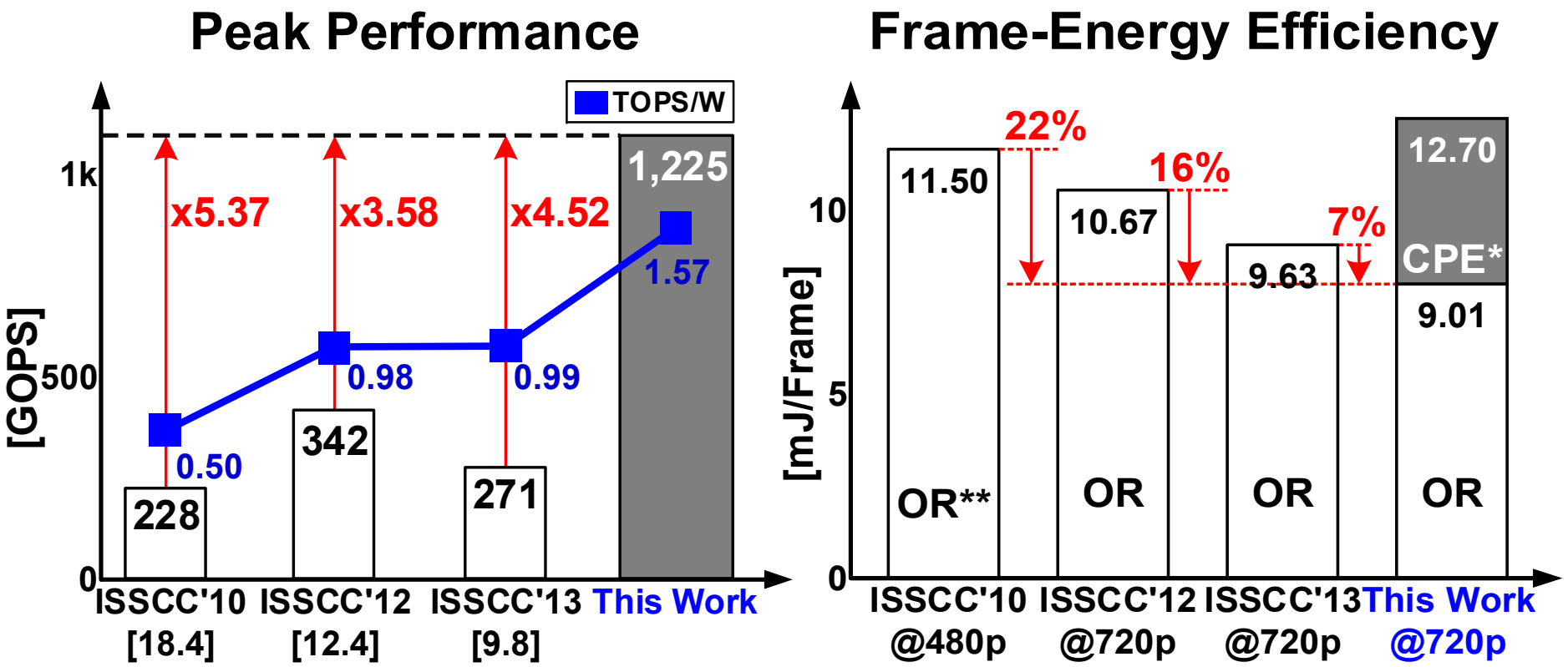


Chip Photograph and Summary



- ❑ 4mm x 8mm by 65nm CMOS Technology
- ❑ 1.22*TOPS* Peak Performance
- ❑ 381*mW* Average / 778*mW* Peak Power Consumption
- ❑ 1.57*TOPS/W* Energy Efficiency for 720p Video

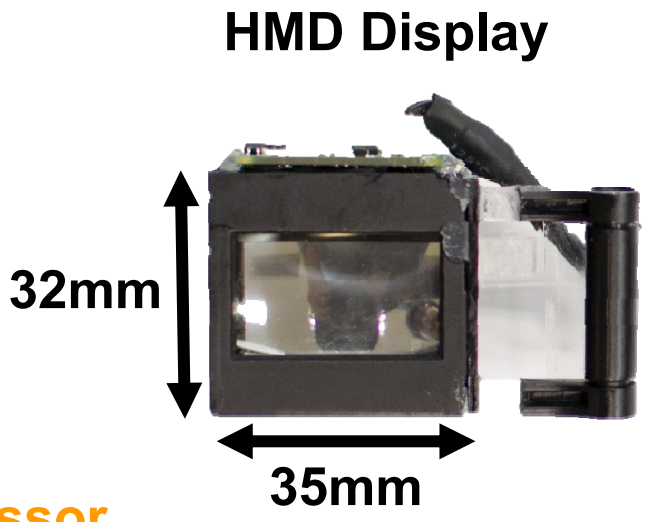
Performance Comparison

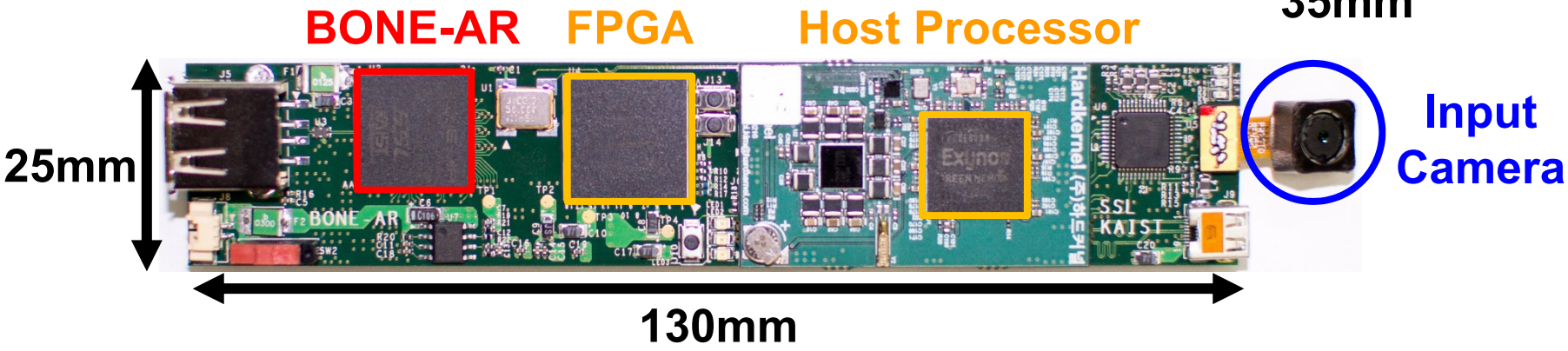


- ❑ **3.58x Peak Performance Increase**
- ❑ **Highest Energy Efficiency 1.57*TOPS/W*, 12.70*mJ/Frame* for 720p Video**

*CPE: Camera Pose Estimation **OR: Object Recognition

K-Glass: HMD System for AR





Demonstration Video

Conclusion

- An HMD Processor for Markerless Augmented Reality

1. High-Throughput (30fps, 1.22TOPS)

➔ Task-level Pipelined Many-SIMD Architecture

2. Low Network Congestion (24.4% ▼)

➔ Neural Network Scheduler for 2D-Mesh NoC

3. Low Energy Consumption (1.57TOPS/W)

➔ Mixed-mode SVM-based Dynamic Resource Management

90nm 20MHz Fully Nonvolatile Microcontroller for Standby-Power-Critical Applications

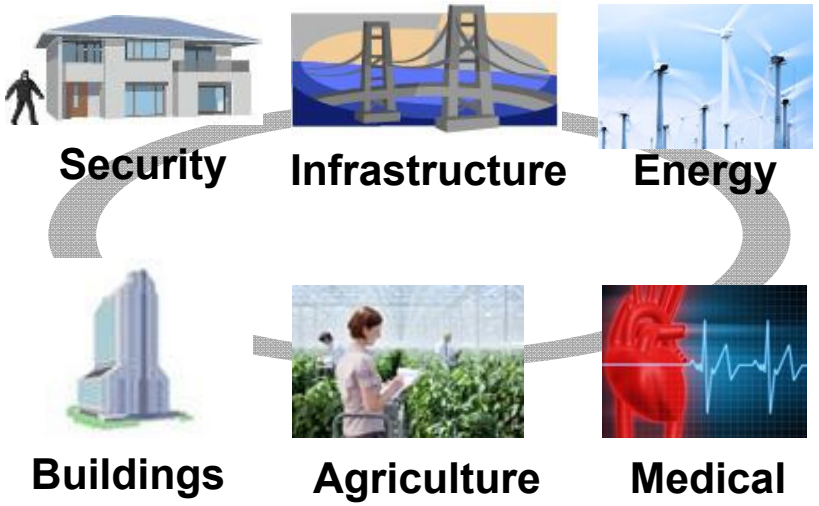
N. Sakimura[1,2], Y. Tsuji[1], R. Nebashi[1], H. Honjo[1,2], A. Morioka[1], K. Ishihara[1], K. Kinoshita[2], S. Fukami[2], S. Miura[1], N. Kasai[2], T. Endoh[2], H. Ohno[2], T. Hanyu[2], and T. Sugibayashi[1]

[1]NEC Corporation, Japan
[2]Tohoku University, Japan

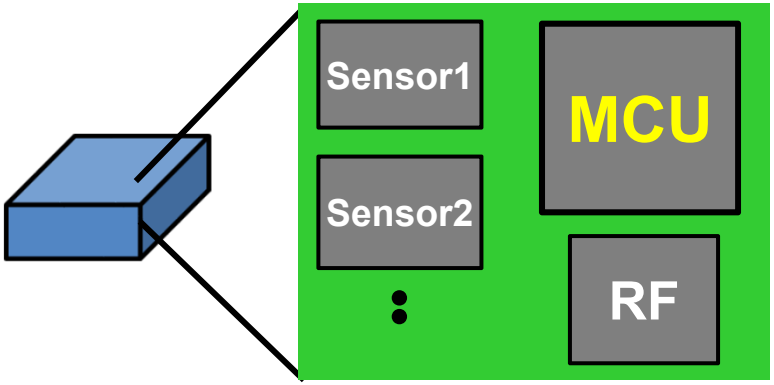
Background (1/2)

- **Increased demand for using "Big Data" collected from wireless sensor networks**
- **Requirements for location- and maintenance-free wireless sensors**
  - Long battery life (e.g. ~10 years), or battery less
  - Data compression and encryption

➔ **Ultra-low-power and high-speed MCU is key component**

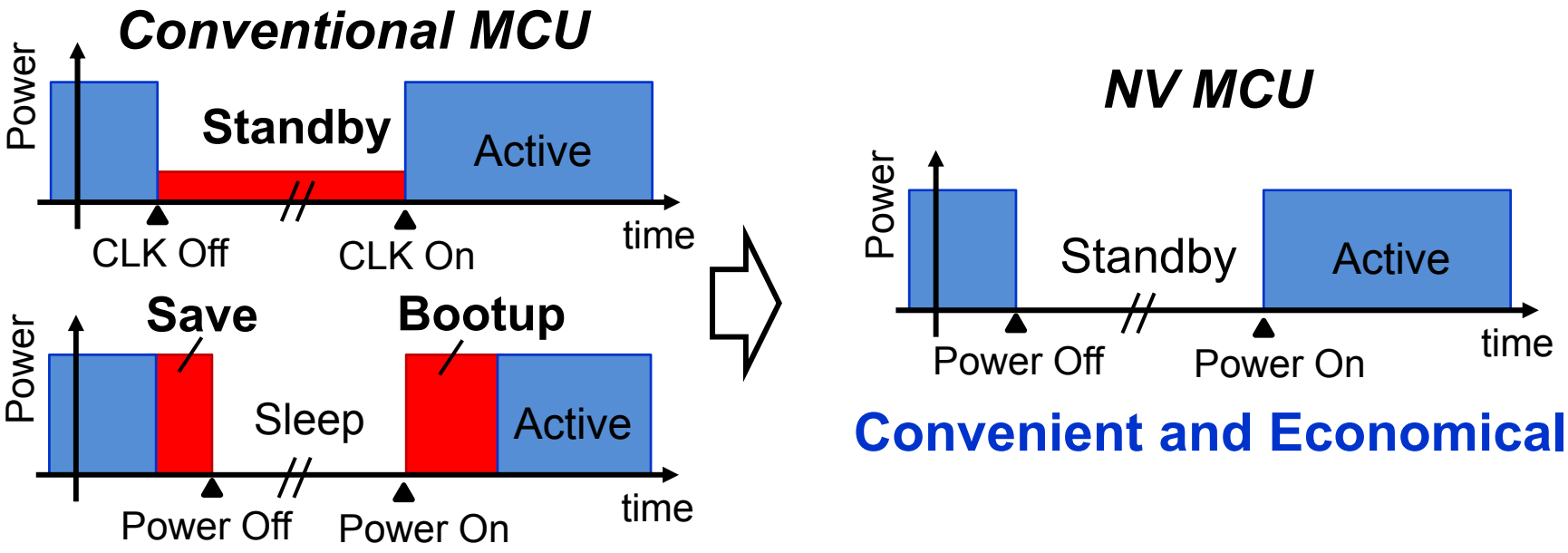


Wireless sensor node



Background (2/2)

- **Discussions on emerging-memory-based nonvolatile (NV) MCU were initiated.**



- **Previous work: 130-nm FeRAM-based fully NV-MCU**

  Bartling et al., ISSCC, 2013.

  - 0.28-μW standby power, 400-ns wakeup time
  - However, 8-MHz operating frequency at 1.5 V…

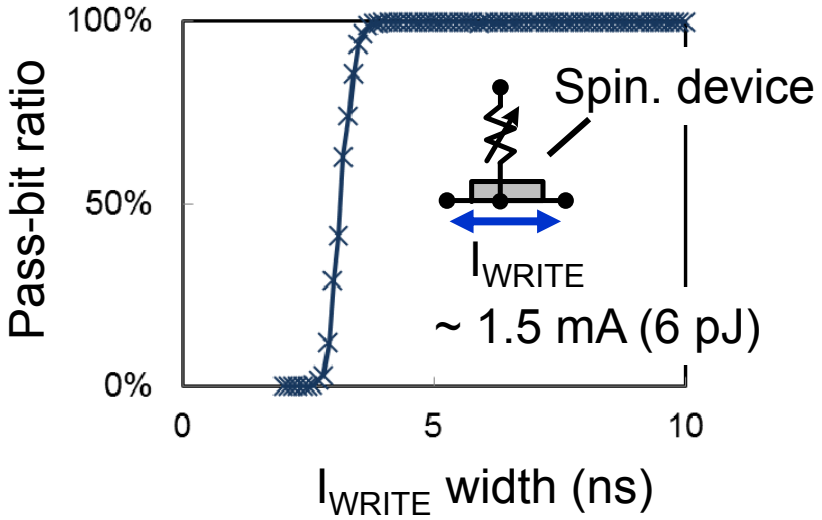
Challenges

Achieving 90-nm fully NV-MCU with SpinRAM

- ~10-ns read/write operations at 1 V

However,

1. **Large leakage: >10x that in 130-nm CMOS**
2. **Large backup energy: ~3x that of FeRAM device**

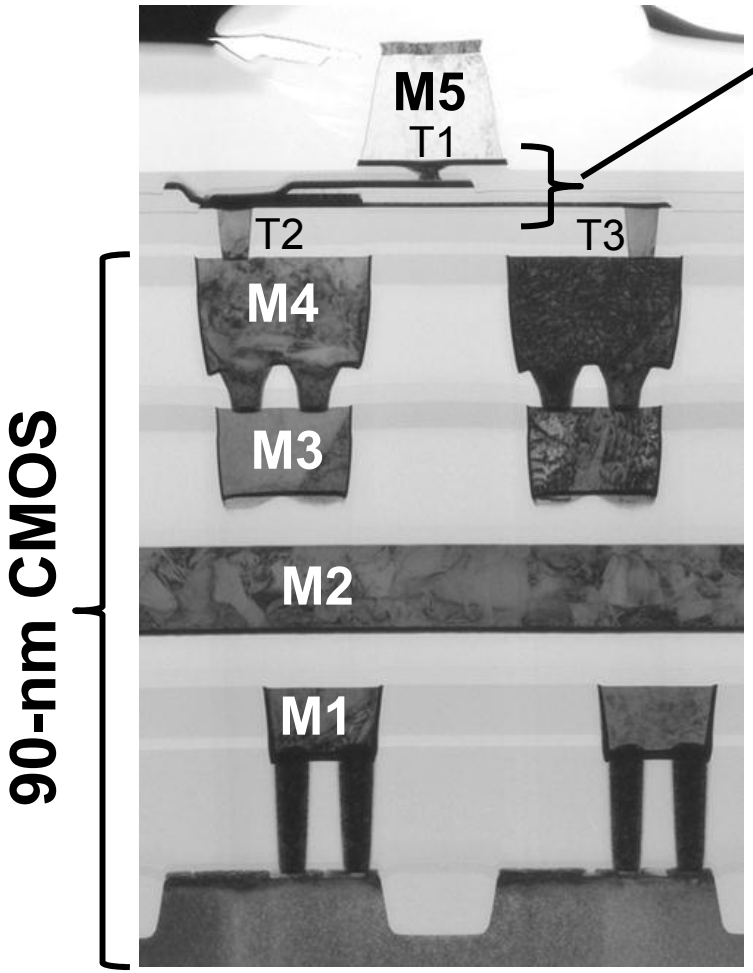


Strategy:

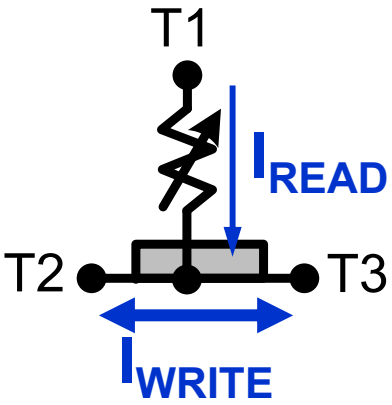
1. **Module-level power gating** for reducing leakage
2. **Optimal backup method** for reducing backup cost in NV logic circuits

90-nm NV-MCU with SpinRAM

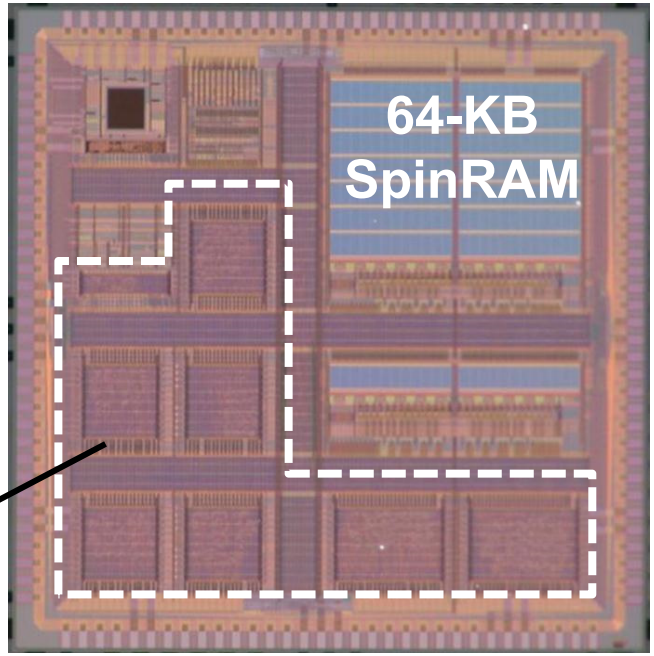
Cross-section



Three-terminal spin. device

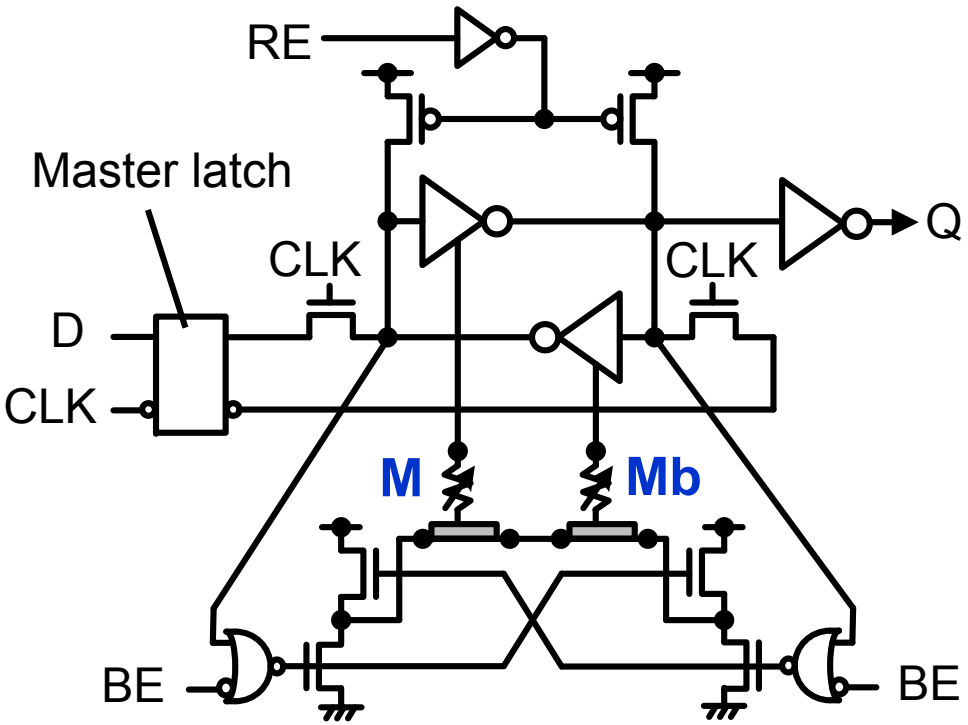


Die photo



NV-logic modules

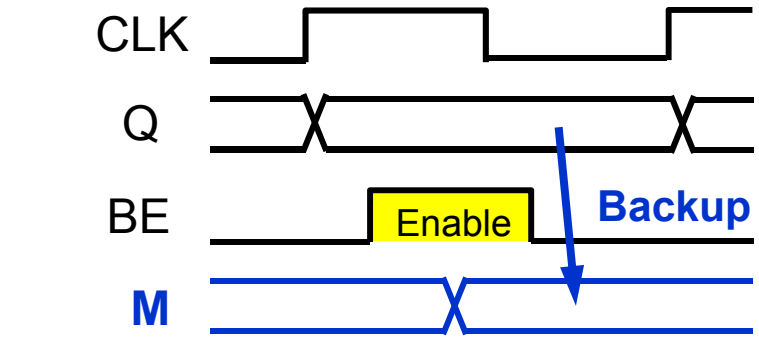
Magnetic F/F (MFF)



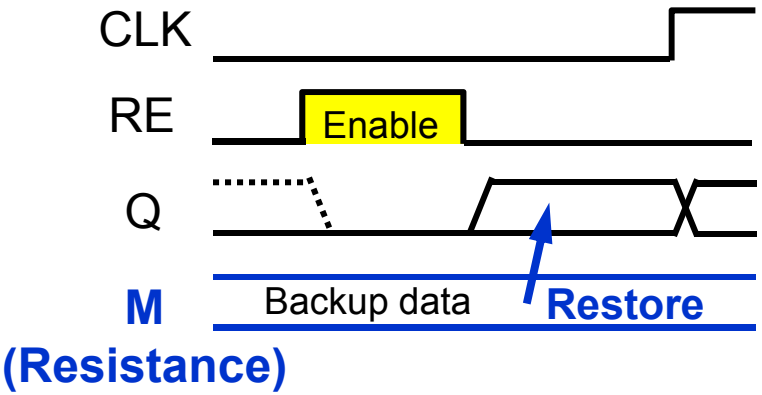
BE: Backup enable input
RE: Restore enable input

N. Sakimura et al., CICC 2008.

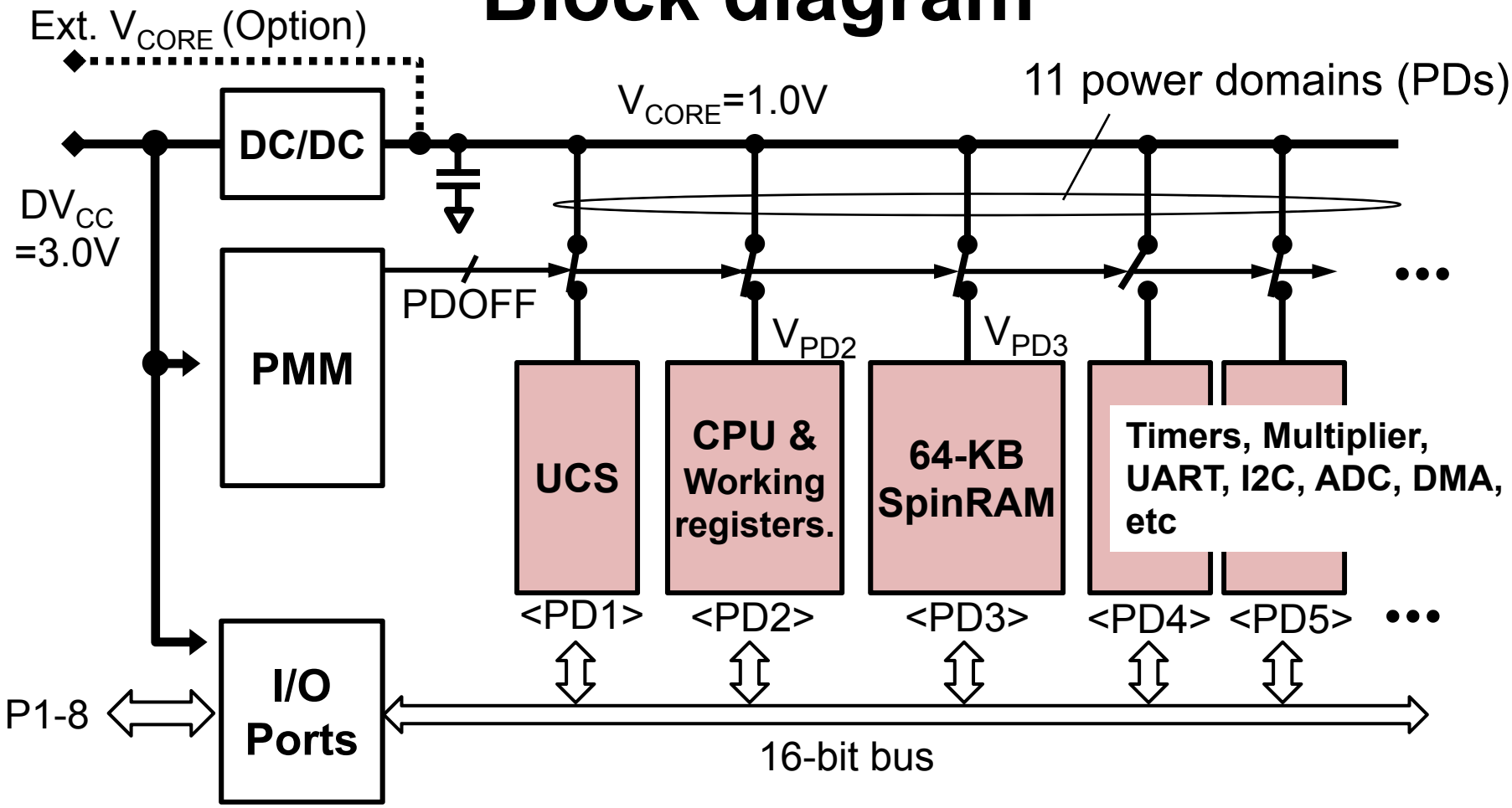
- **Backup operation**



- **Restore operation**



Block diagram

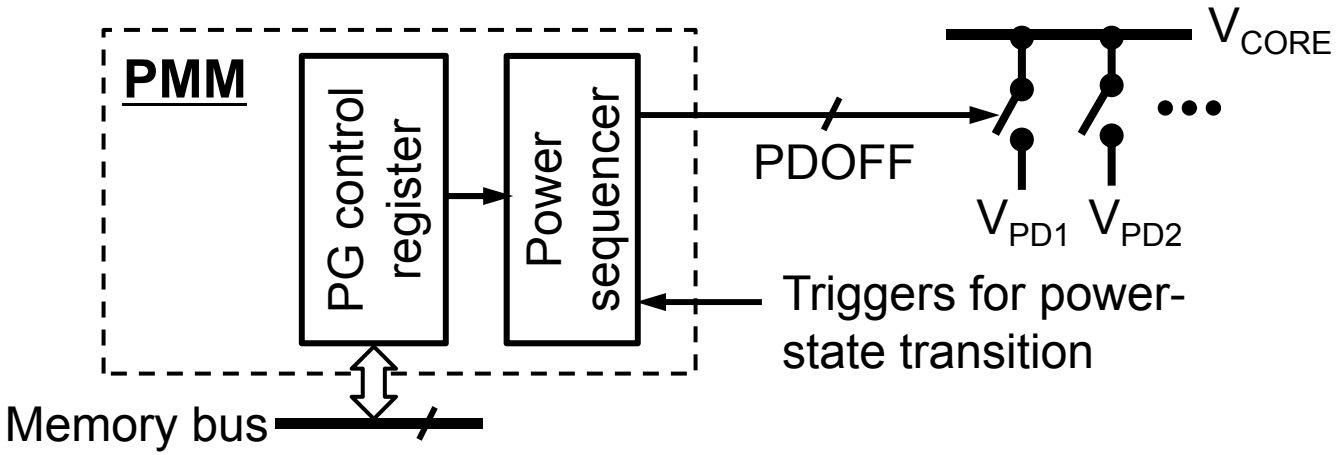


PMM: Power Management Module
UCS: Unified Clock System
UART: Universal Asynchronous Receiver Transmitter
I2C: Inter-Integrated Circuit
ADC: Analog-to-Digital Converter
DMA: Direct Memory Access

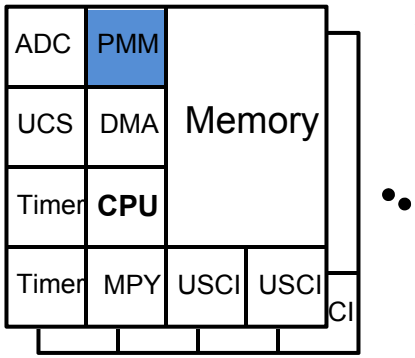
Always On

NV Modules

Module-level power gating



Almost Off (Standby)

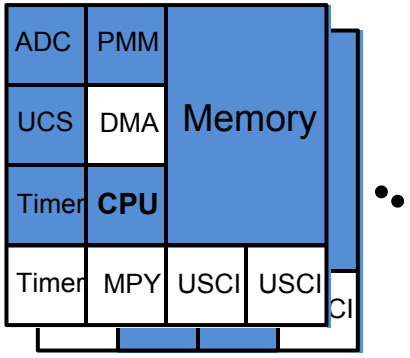


Zero leakage (Min.)
(1.6-µW standby power)

Timer/Ext. wakeup w/o bootup code

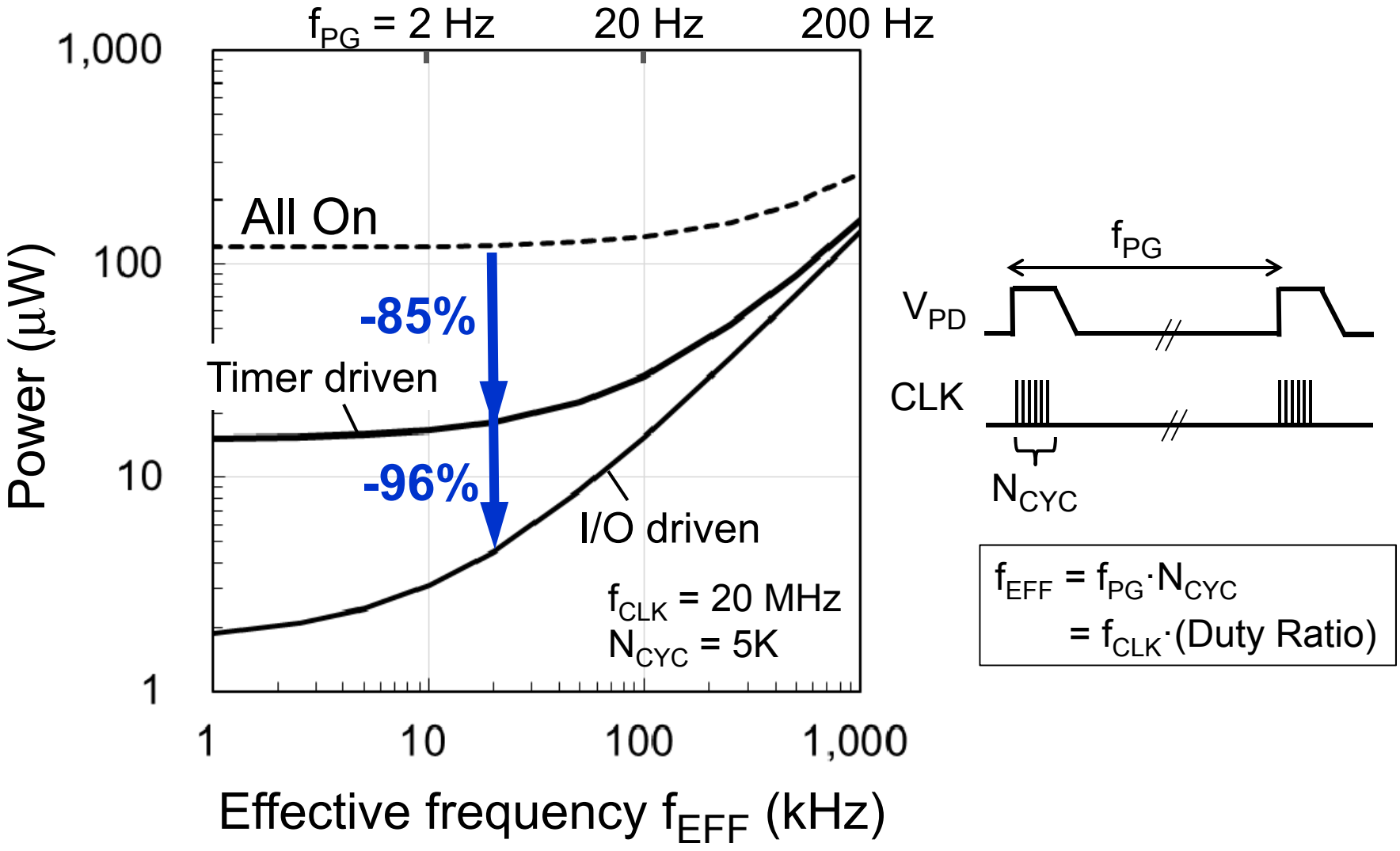
CPU request

Partially On (Active)

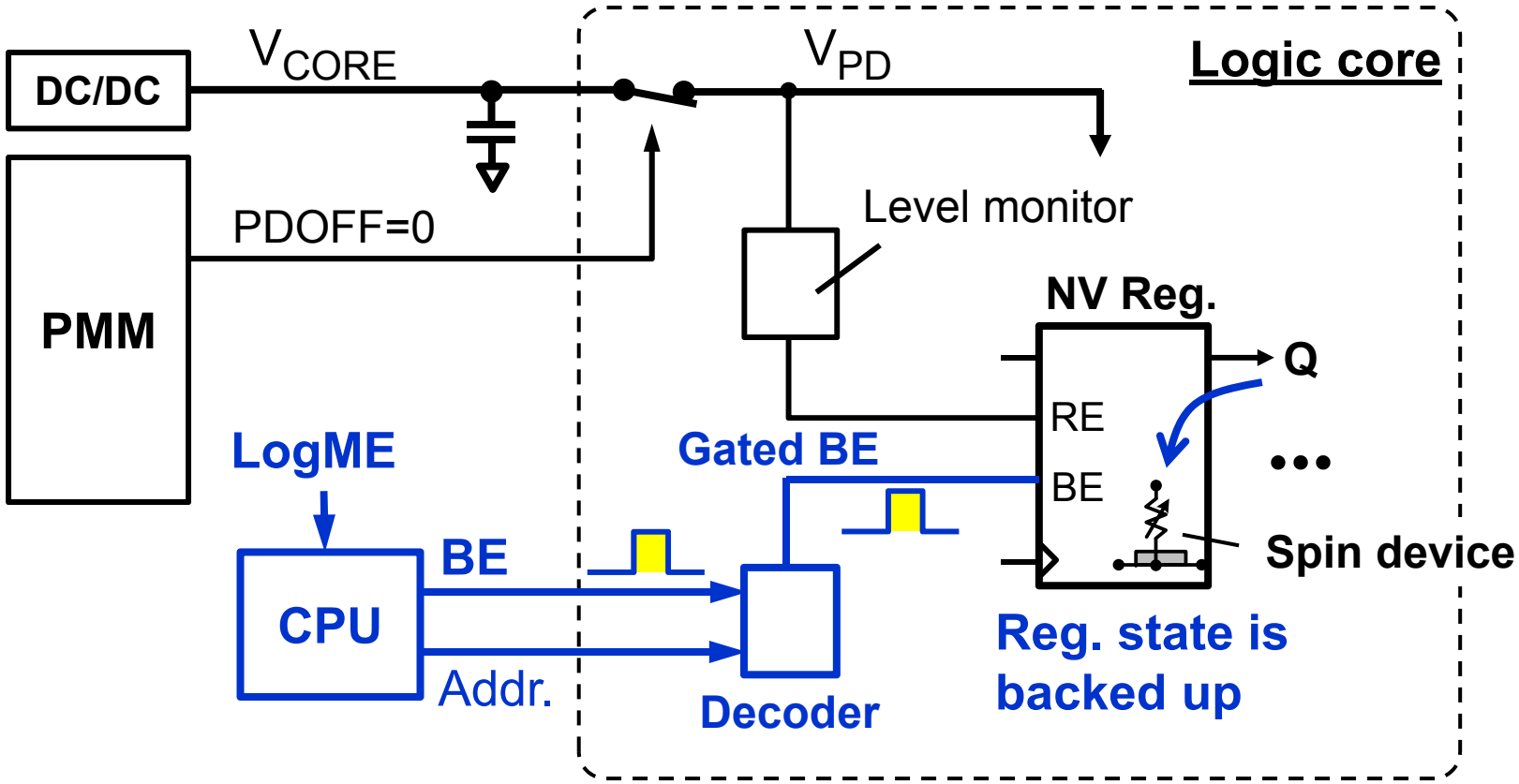


<117-µW leakage (Max.)

Estimated power dissipation



NV-reg. backup



- Word-level backup for peak-current suppression
- Backup control for its power reduction
  - by software with “SAVE” instruction
  - by hardware in logging mode (LogME=1)

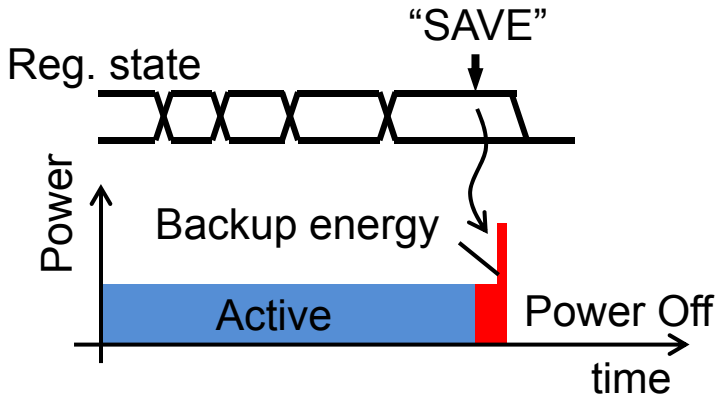
Backup power reduction (1/2)

Software-controlled backup

- 530 pJ/word backup energy
- It can prevent from overwriting by backing up just before power off.

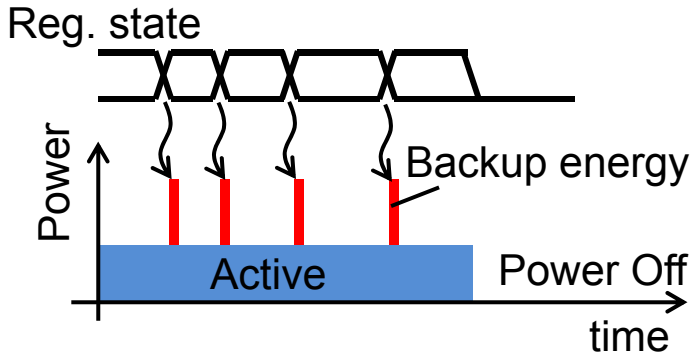
"SAVE" instruction

```
SAVE   R1    : work.Reg.(1 cycles)
SAVE   &Add  : absolute (3 cycles)
SAVE   X(R5): index (4 cycles)
```

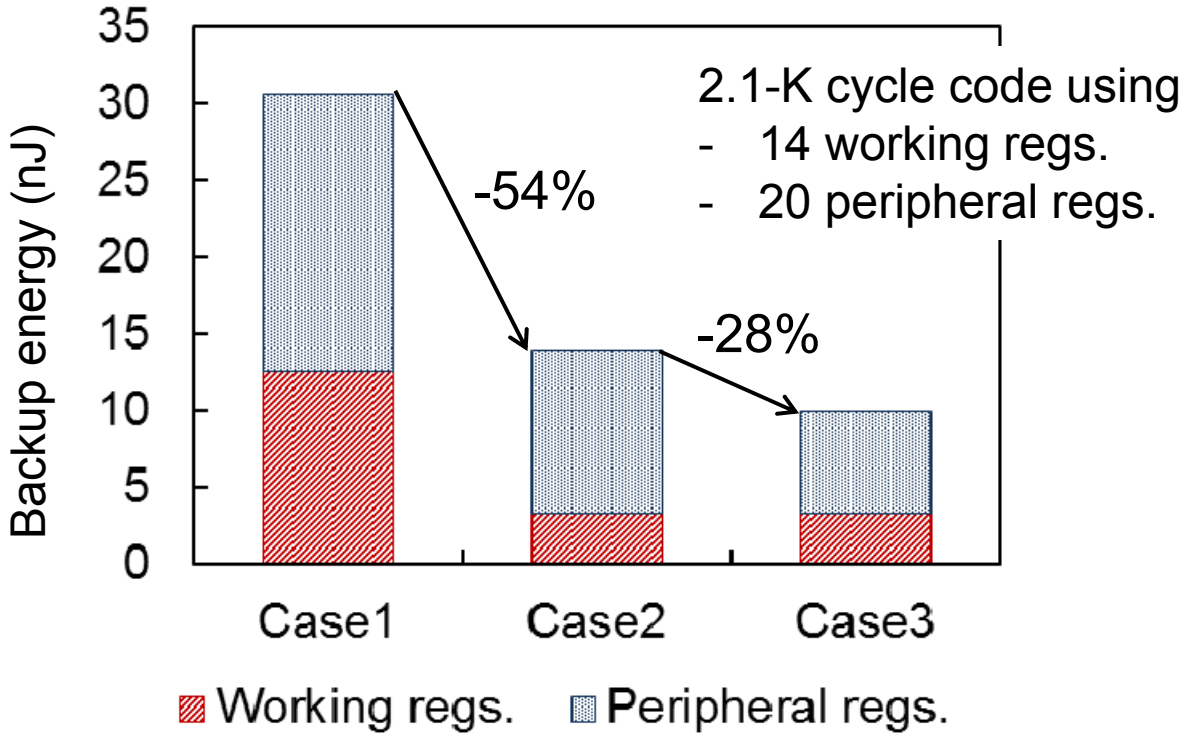


Logging backup (Not involving SW)

- 96 pJ/word backup energy
- However, it increases overwriting.

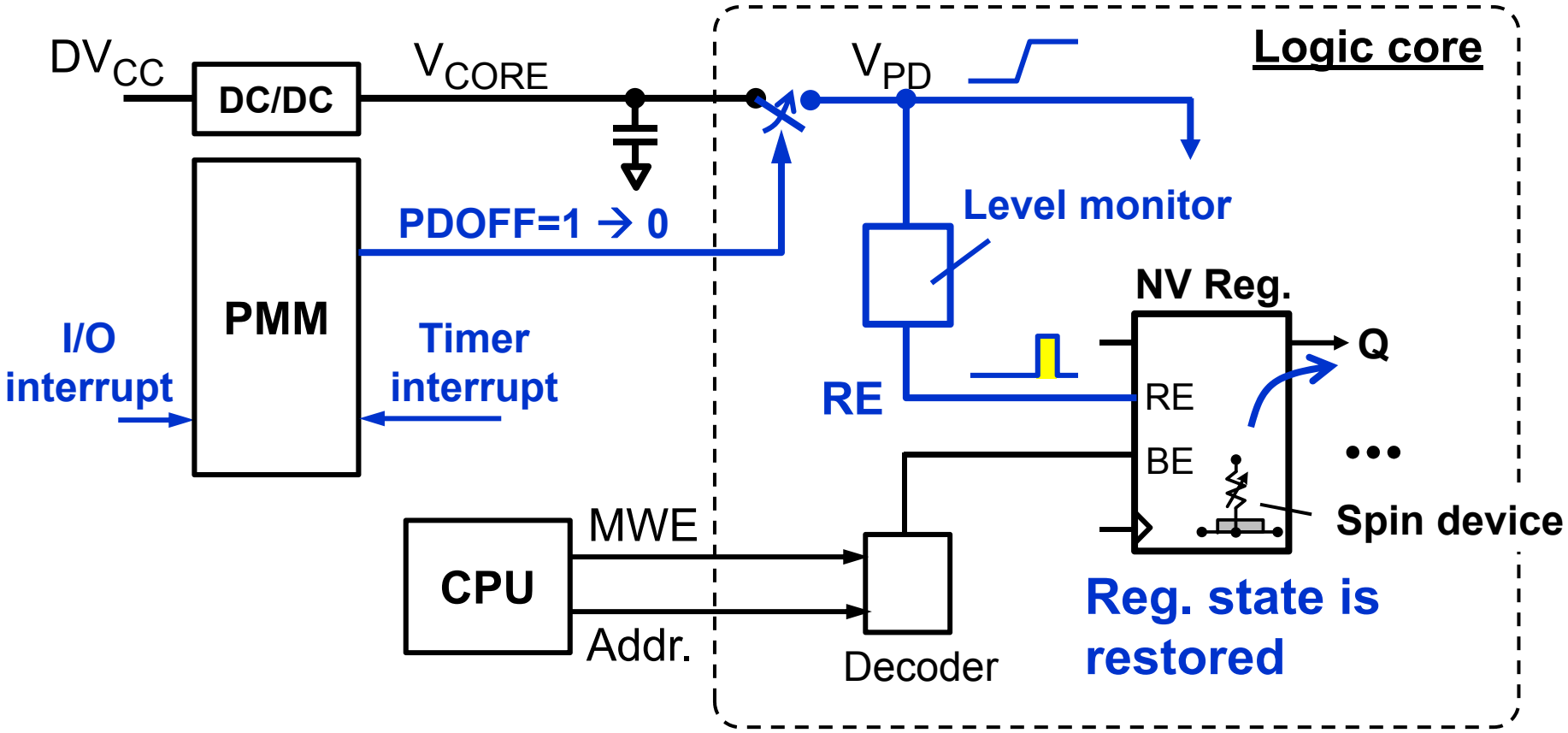


Backup power reduction (2/2)



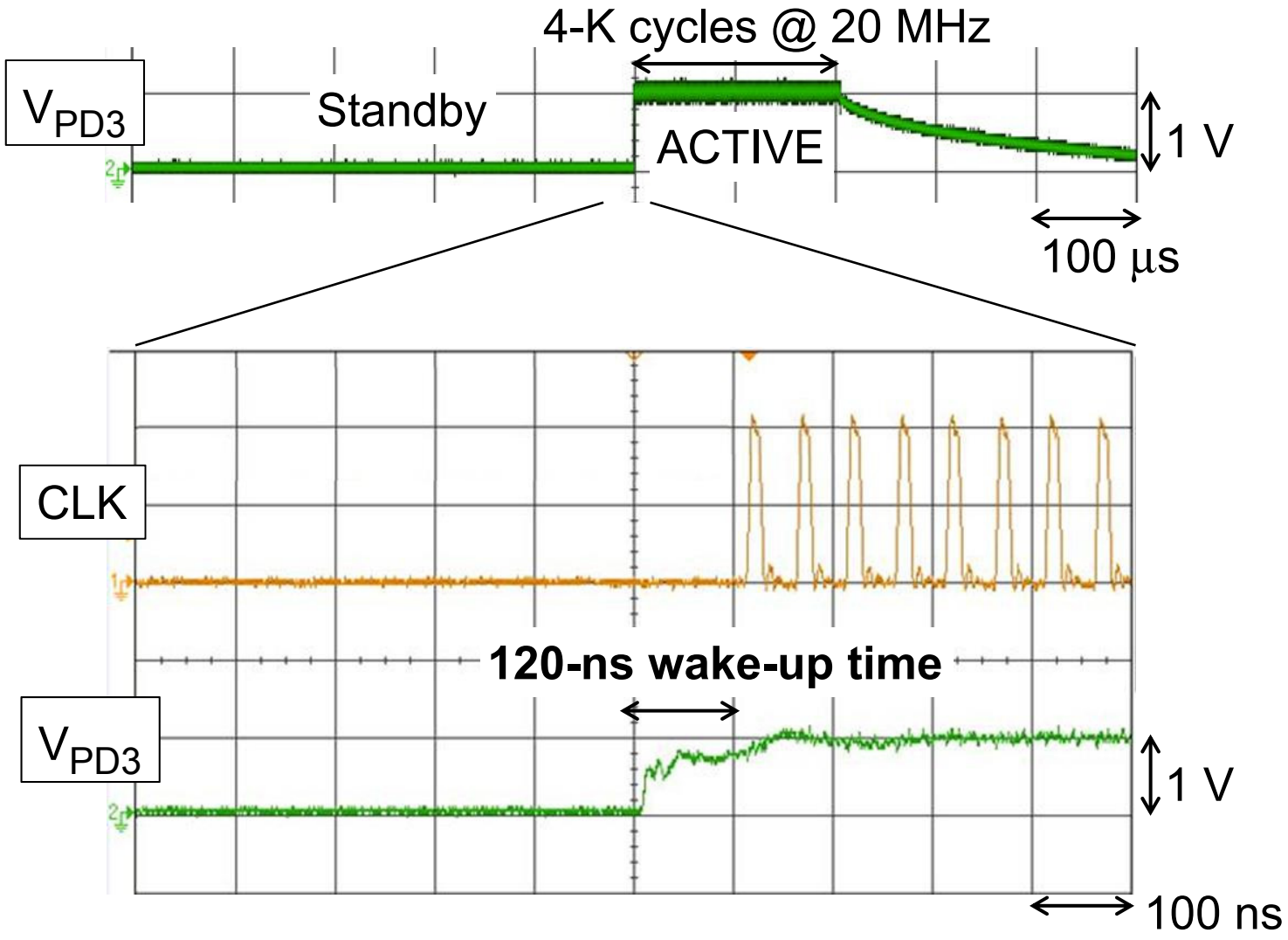
	How to store data	Working regs.	Peripheral regs.
Case 1	Transfer to memory macro	S/W control	
Case 2	Using NV register	S/W control	
Case 3		**S/W control**	**Logging**

NV-reg. restore

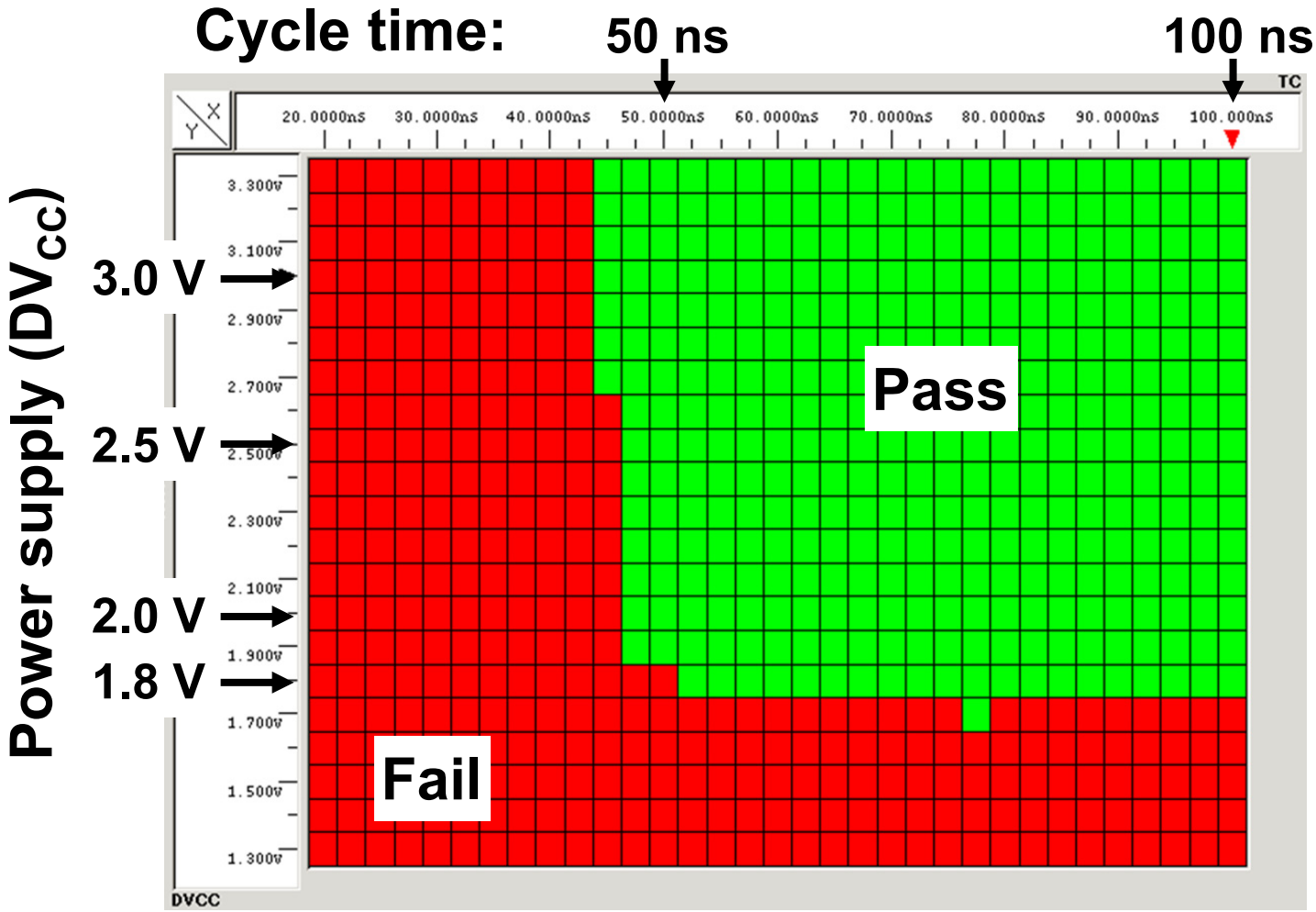


- Module-level restoring
- Restore triggers
  - Timer or I/O request
  - Software request

Wakeup time



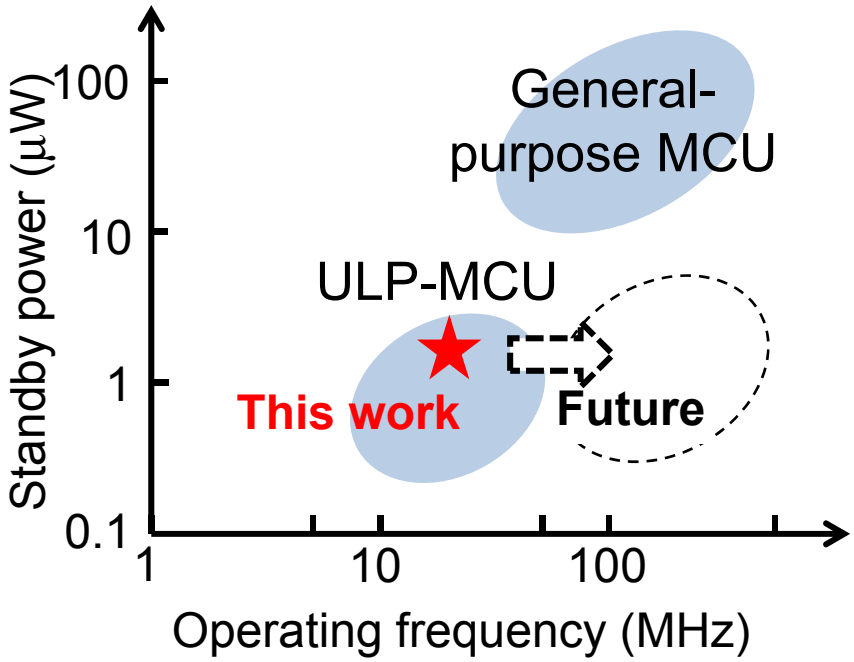
Operation range



Conclusions

90-nm 20-MHz fully NV-MCU with three-terminal SpinRAM

- Module-level power gating for reducing leakage
- Optimal backup control for NV logic



Acknowledgements

This work is supported by JSPS through FIRST program

A 0.74V 200µW Multi-Standard Transceiver Digital Baseband in 40 nm LP CMOS for 2.4 GHz Bluetooth Smart / Zigbee / IEEE 802.15.6 Personal Area Networks

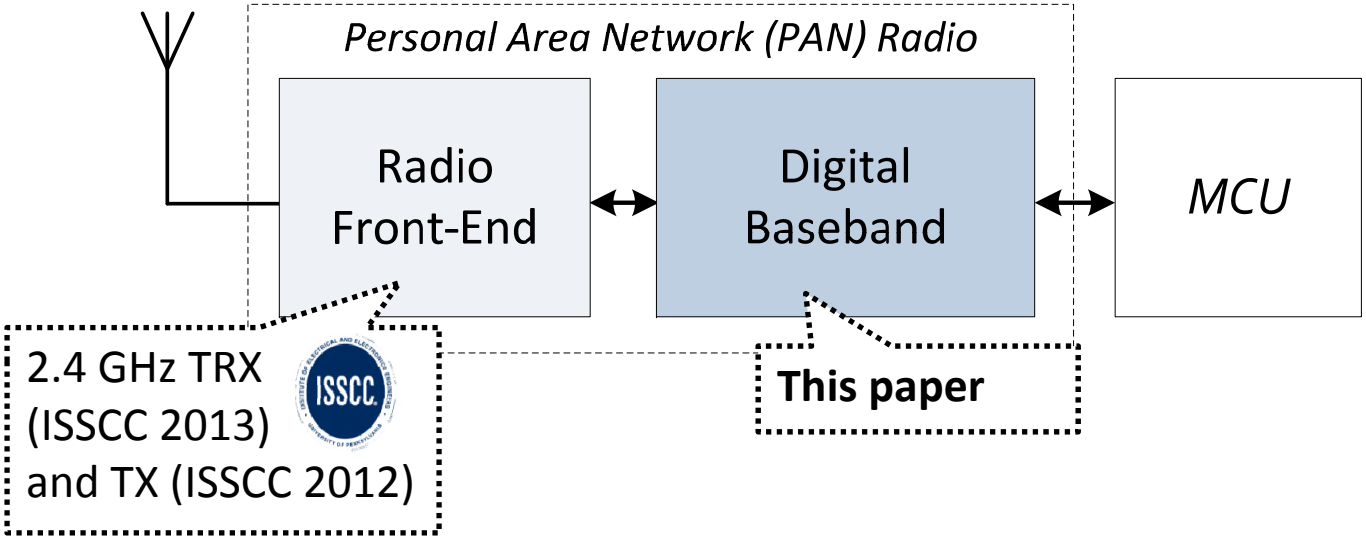
Christian Bachmann, Gert-Jan van Schaik, Benjamin Busze, Mario Konijnenburg, Yan Zhang, Jan Stuyt, Maryam Ashouei, Guido Dolmans, Tobias Gemmeke, Harmke de Groot

imec - Holst Centre, Eindhoven, The Netherlands

Outline

- Scope
- Architecture
- Implementation
- Results
- Conclusion

Personal Area Networks Radio



- Design goals
  - Multi-standard functionality
  - Low power: < 6 mW (TX/RX)
    - i.e., DBB < 0.5 mW

Bluetooth Smart / Low Energy (BTLE)

ZigBee IEEE 802.15.4

IEEE 802.15.6 (MBAN)

Multi-standard Digital Baseband: Scope

MAC & SW SUPPORT

- Register interface towards SW protocol control and data source / sink (MCU)
- MAC protocol timing support

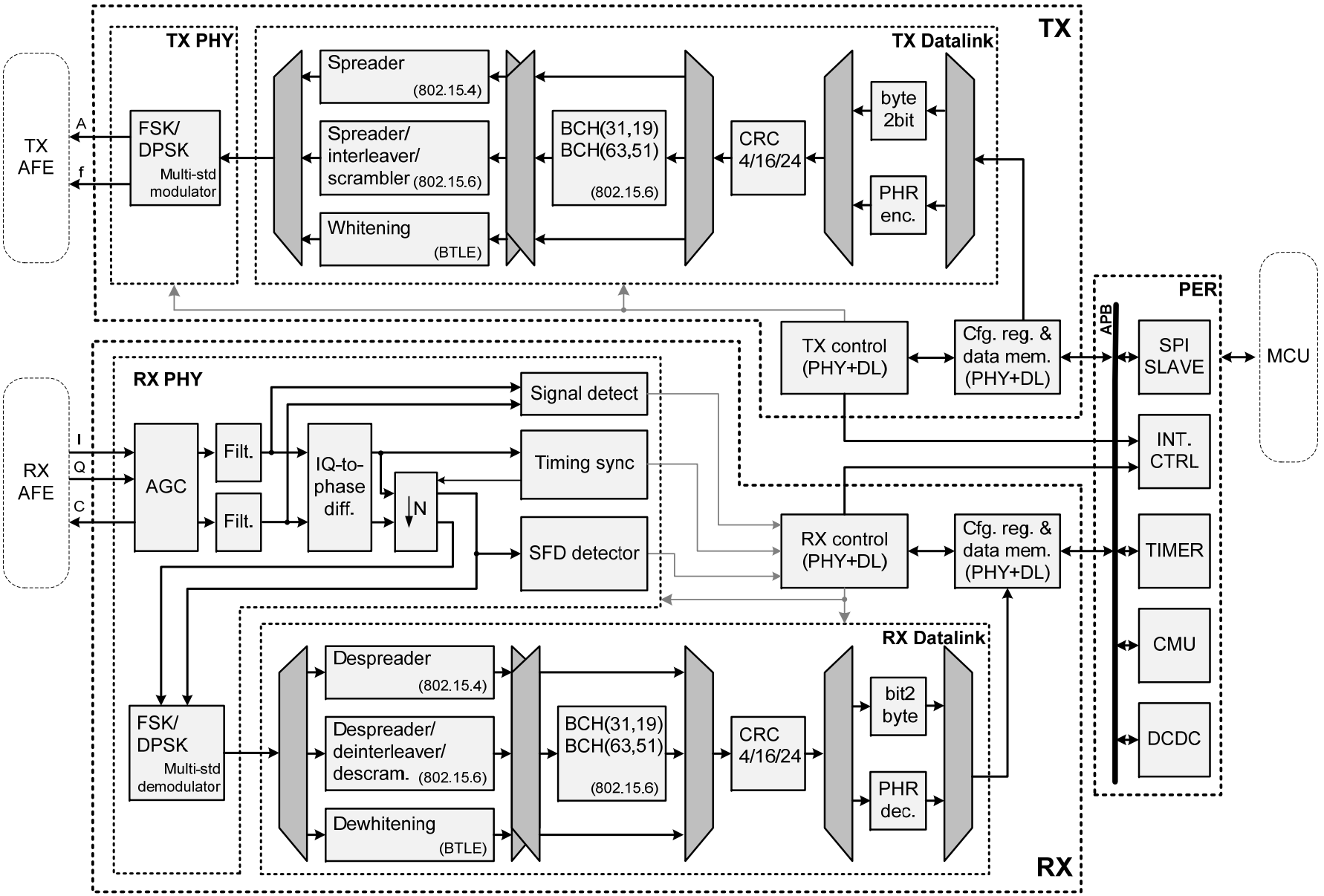
LOWER DATA LINK LAYER

- PHR / PSDU en- & decoding (all stds.)
- Error detection: CRCs (all stds.)
- Error correction: BCH FEC (802.15.6)

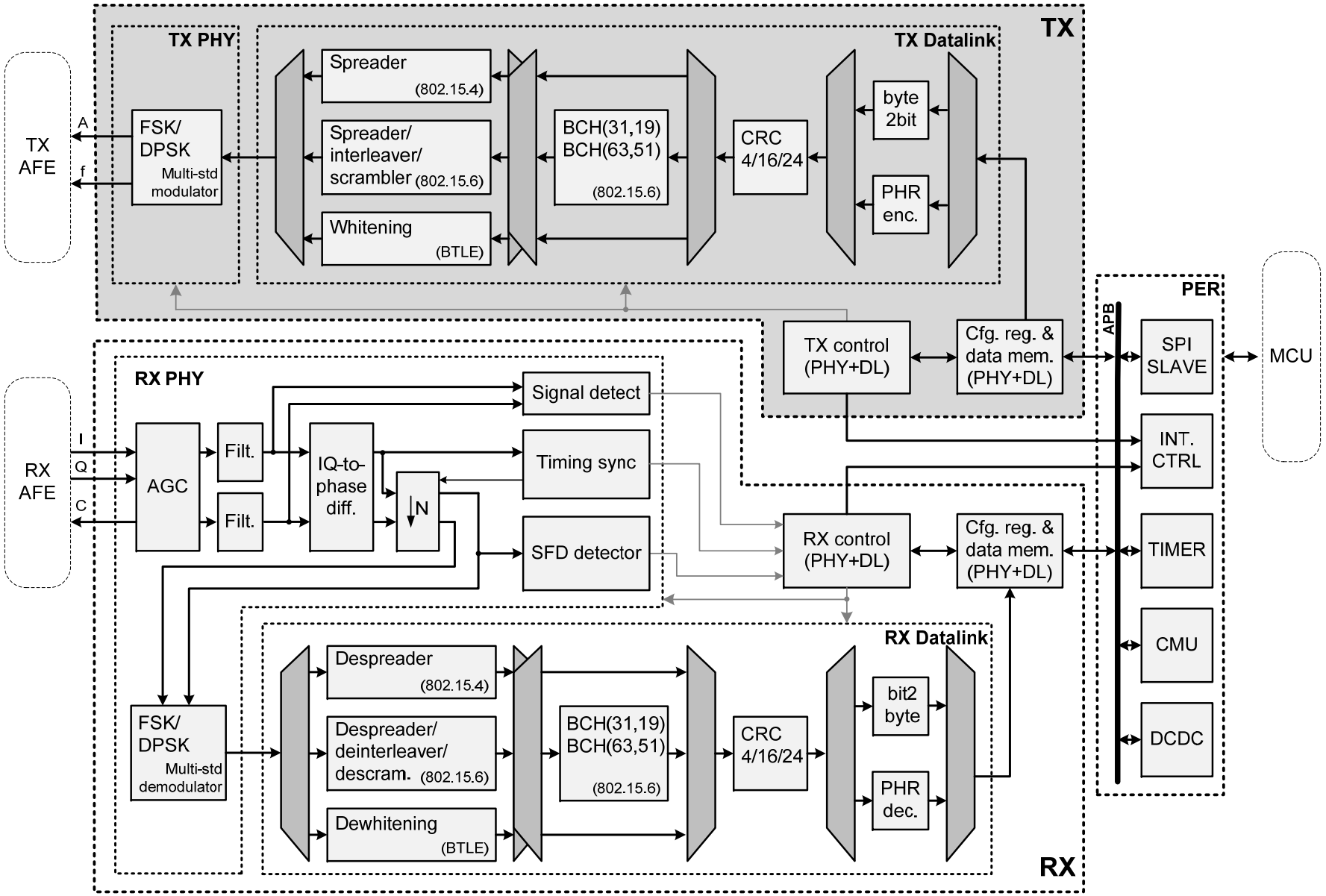
PHYSICAL LAYER

- Modulation (A/f) & demodulation (IQ)
  - GFSK, OQPSK, DB/QPSK schemes
- Automated Gain Control (AGC)
- Synchronization (symbol & frame)

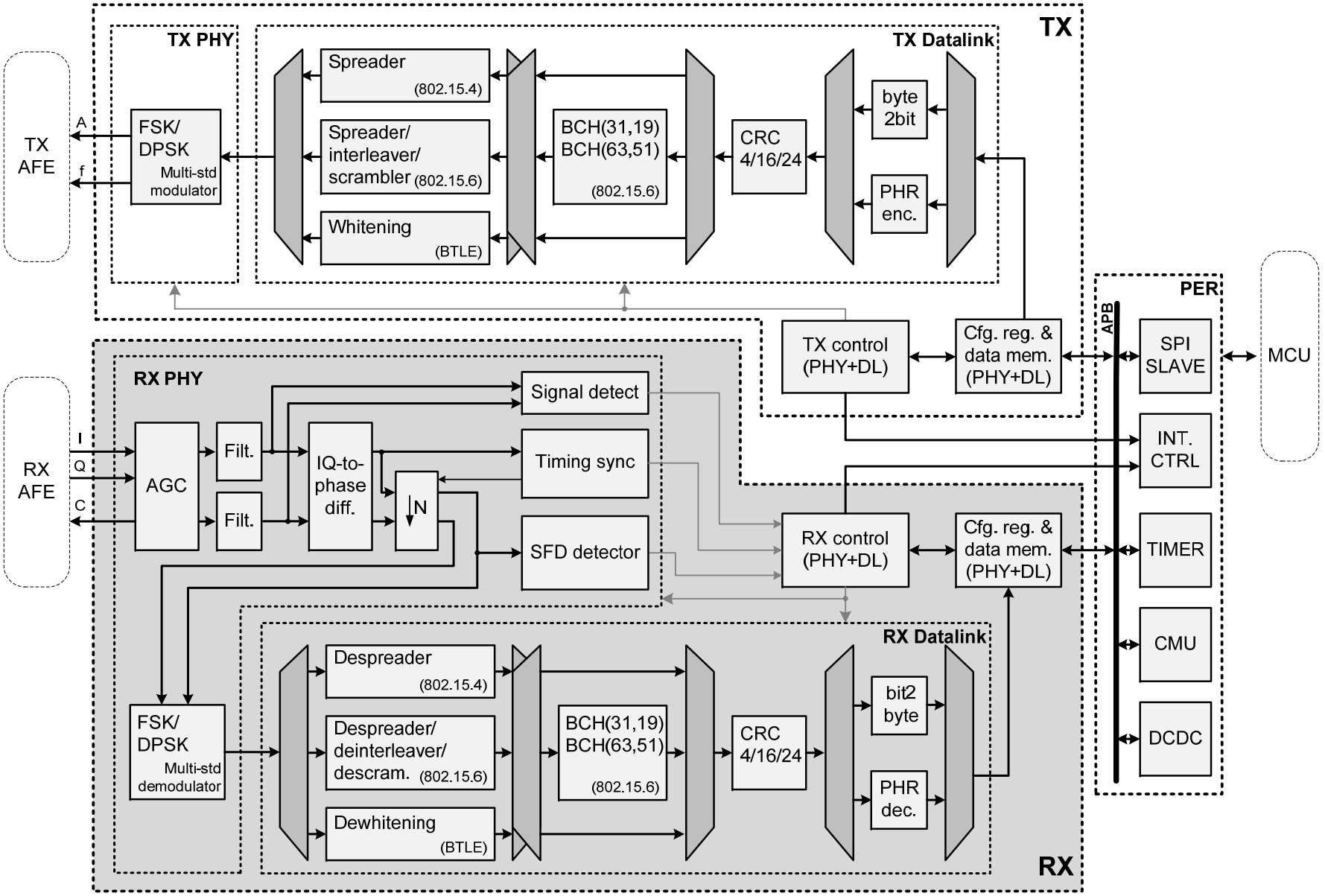
Multi-standard DBB Architecture



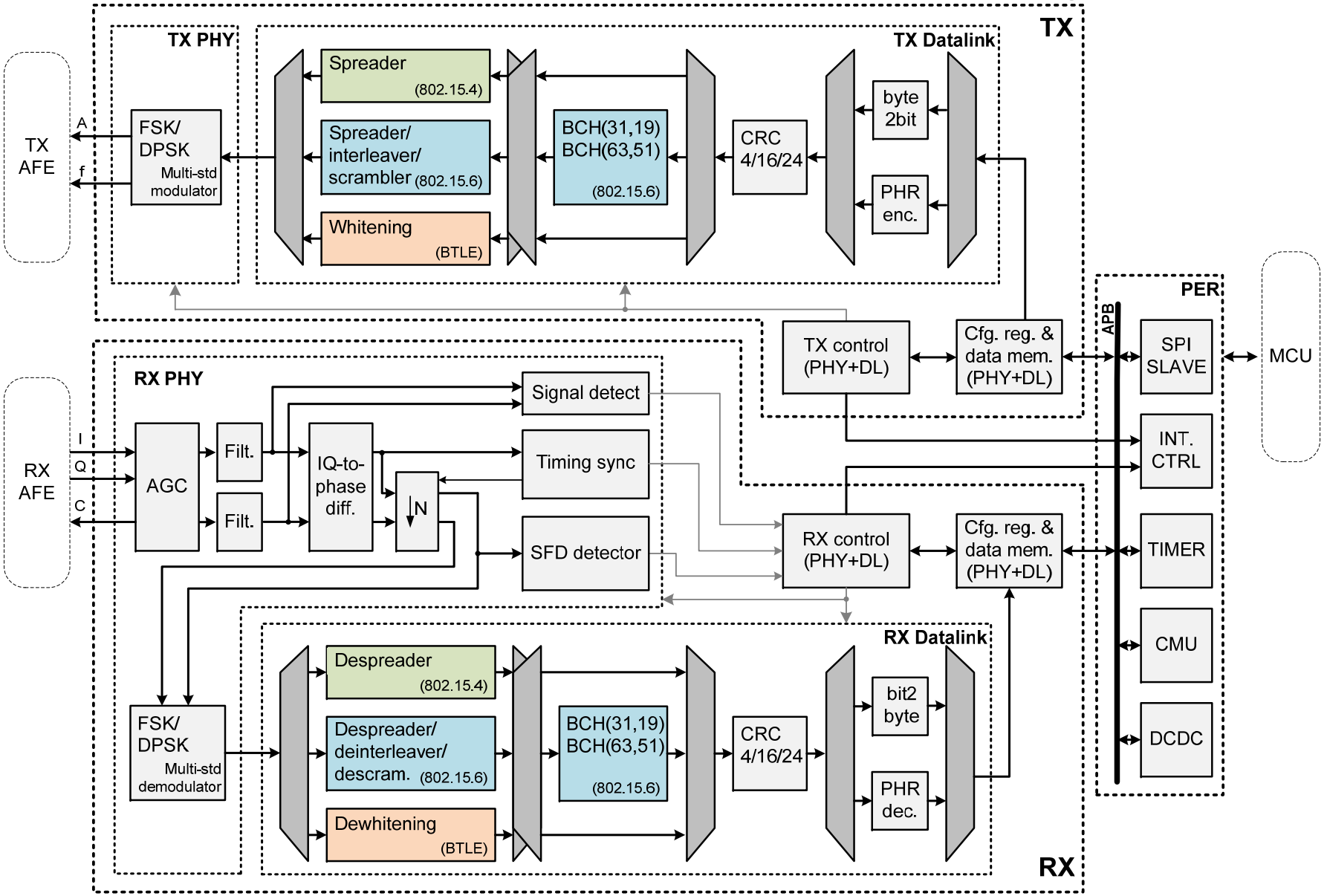
DBB Architecture: TX



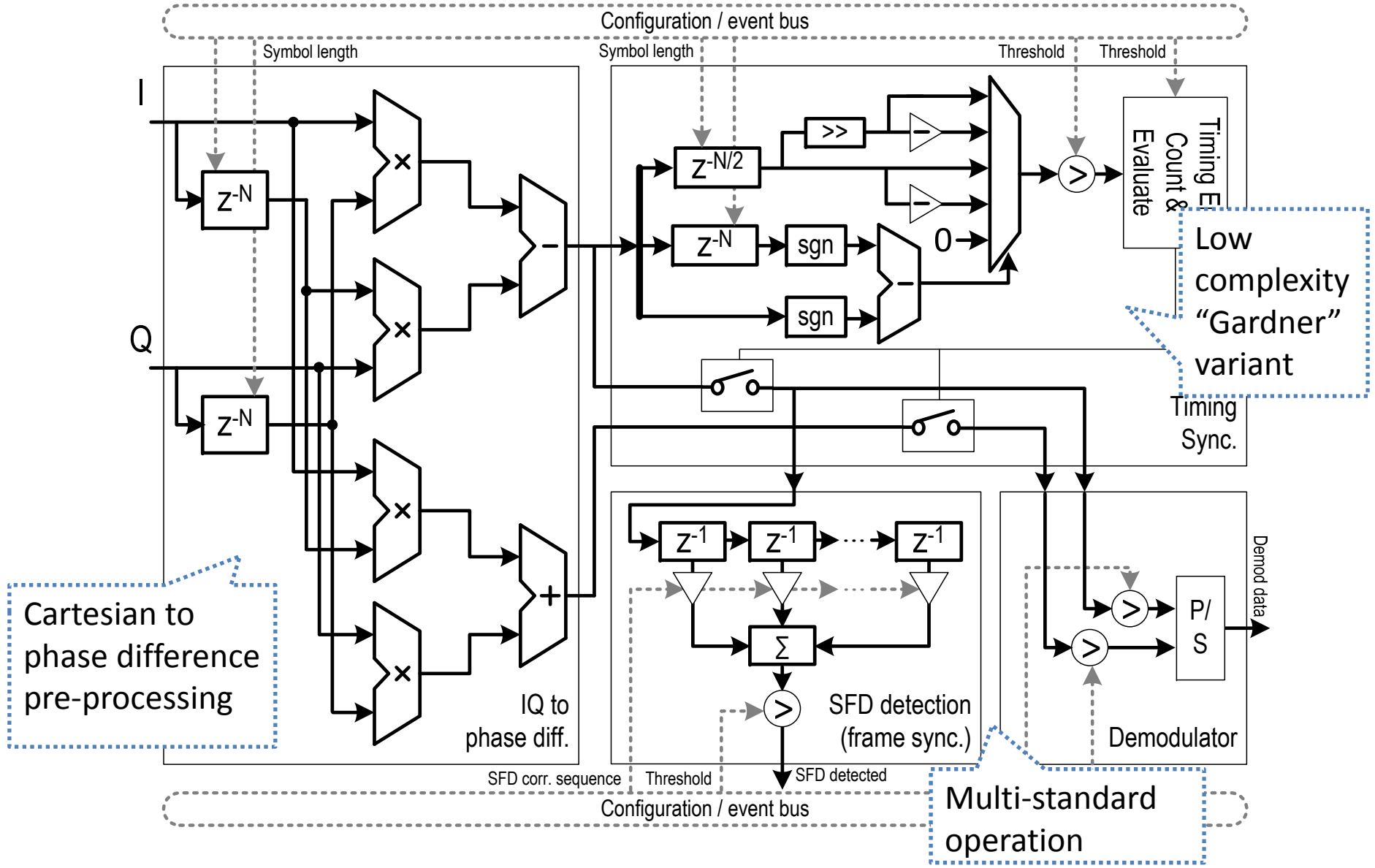
DBB Architecture: RX



DBB Architecture: Standard-specific

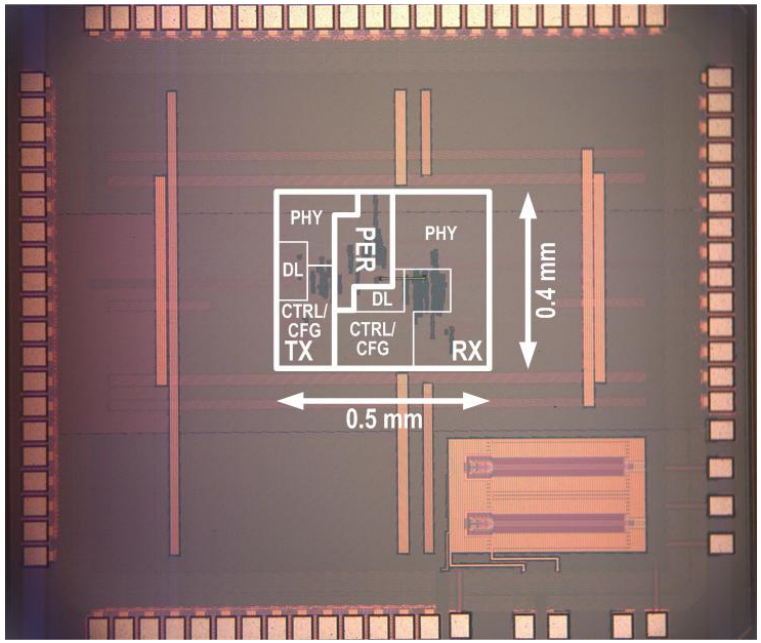


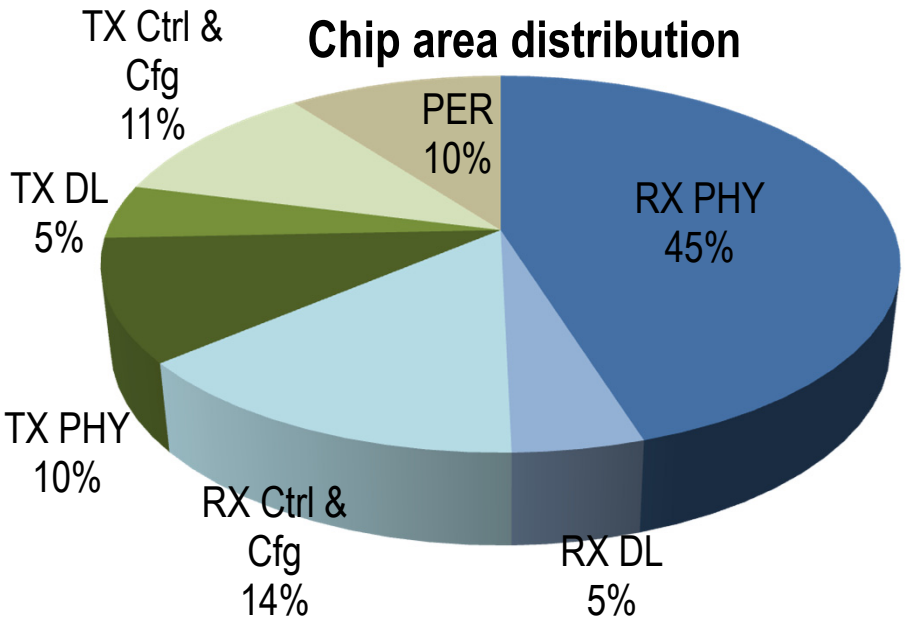
Optimized Implementation: RX PHY

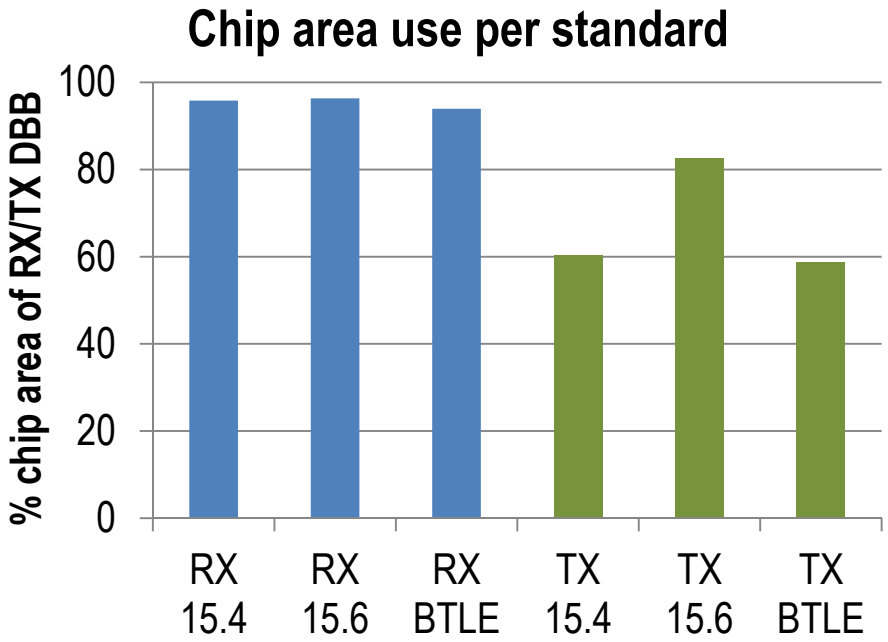


Chip Implementation

- TSMC 40nm LP CMOS
- Core area 0.2 mm^2







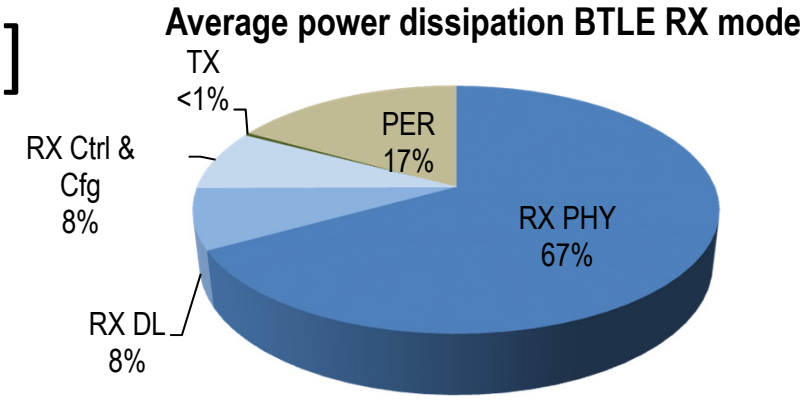
Power Consumption

- DBB measured active mode power @ 0.74 [V]

Use case	Bluetooth Smart (BTLE)	IEEE 802.15.4	IEEE 802.15.6	[3] IEEE 802.15.4	[4] IEEE 802.15.4
RX active mode [µW]	< 180	< 200	< 140	6660 (~ 250)*	5760 (~ 216)*
TX active mode [µW]	< 80	< 80	< 60	5040 (~ 189)*	4140 (~ 155)*

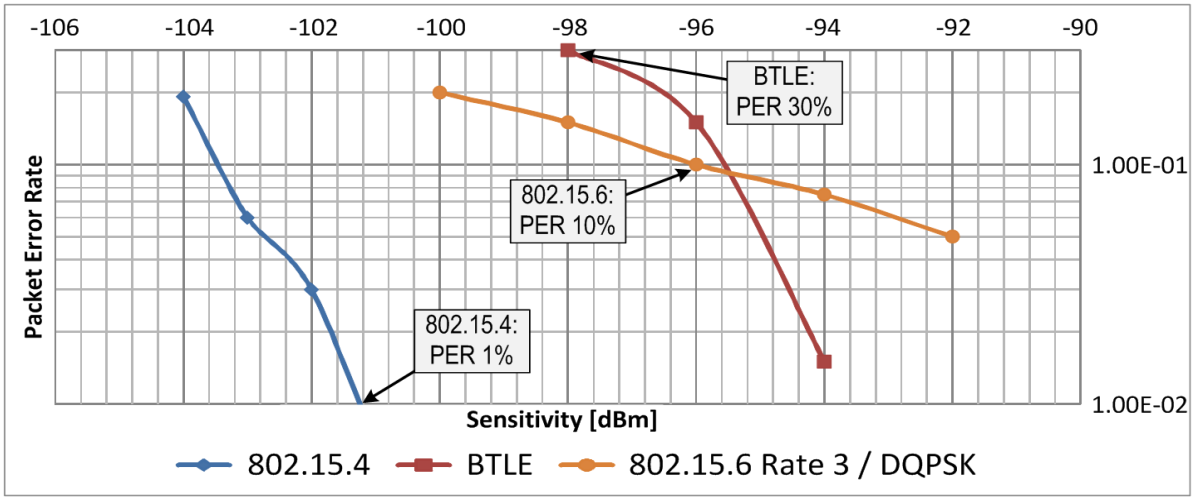
* Considering geometric (0.18µm to 40nm) and supply voltage (1.8V to 0.74V) scaling.

- Including RF front-end [1]
  - RX < 4 mW
  - TX < 5.5 mW



Evaluation

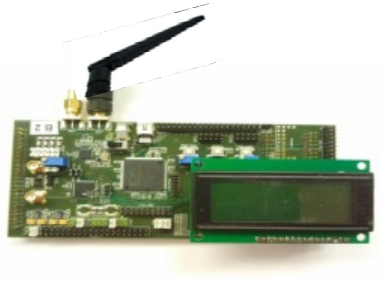
- Evaluation of front-end [1] and DBB[1)]



[1)] Standard requirements: > -70 dBm, (BTLE) > -85 dBm (802.15.4), > -83 dBm (802.15.6)

- Demo: Heart rate calculation[2)] and transmission via Bluetooth Smart (BTLE) link to COTS device

ISSCC 2014
Demo Session



Smart / Low Energy (BTLE)

[2)] *N. van Helleputte, et al., "A Multi-Parameter Signal-Acquisition SoC for Connected Personal Health Applications", ISSCC 2014*

Conclusion

- Multi-standard DBB for PAN radio
  - Physical and lower datalink processing
  - Bluetooth Smart, IEEE 802.15.4 and 802.15.6
- Low power consumption & small area
  - Optimized algorithmic complexity
  - Resource sharing across standards
  - Power < 200 / 80 uW (RX/TX)
  - Area < 0.2 mm2 @ 40 nm

A 105GOPS 36mm2 Heterogeneous SDR MPSoC with Energy-Aware Dynamic Scheduling and Iterative Detection-Decoding for 4G in 65nm CMOS

Benedikt Noethen, Oliver Arnold, Esther Perez Adeva, Tobias Seifert, Erik Fischer, Steffen Kunze, Emil Matúš, Gerhard Fettweis, Holger Eisenreich, Georg Ellguth, Stephan Hartmann, Sebastian Höppner, Stefan Schiefer, Jens-Uwe Schlüßler, Stefan Scholze, Dennis Walter, René Schüffny

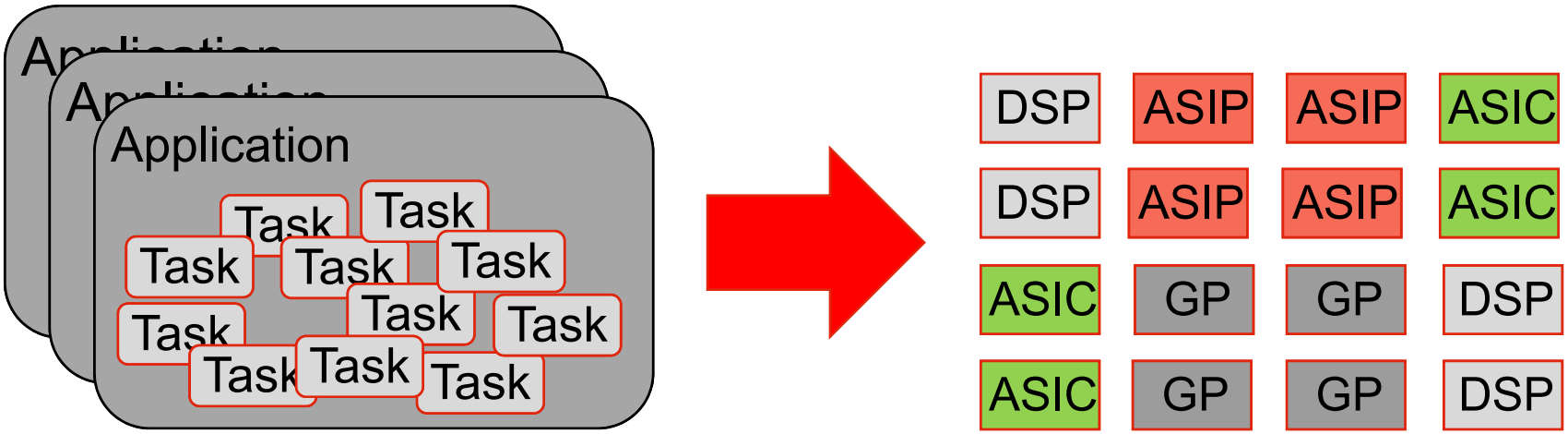
Technische Universität Dresden, Germany

SPONSORED BY THE
Federal Ministry of Education and Research

Outline

- Motivation
- Tomahawk Architecture Framework
- Chip Implementation
  - CoreManager
  - Power Management
  - Processing Elements
  - Iterative Detection-Decoding
- Conclusion

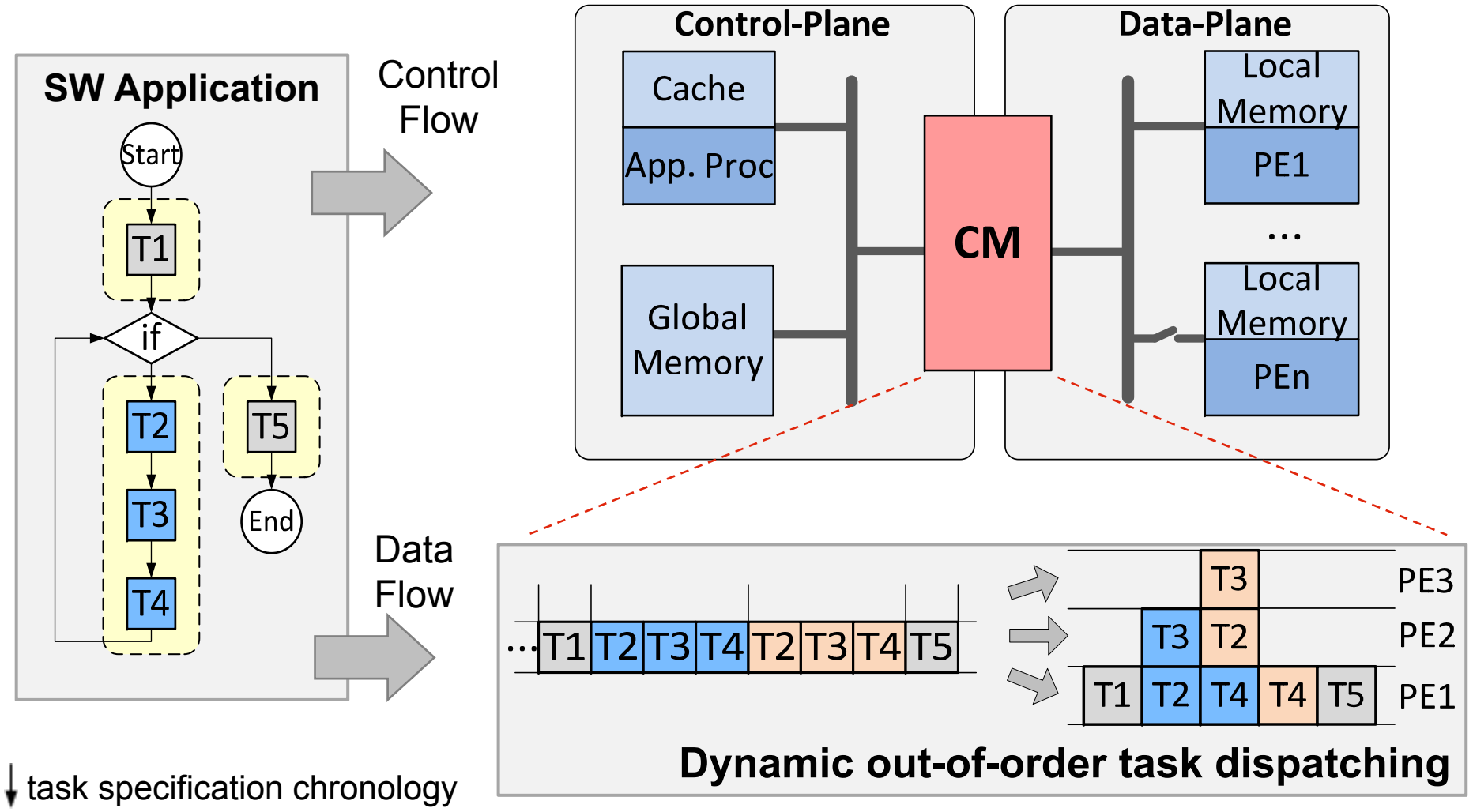
Motivation



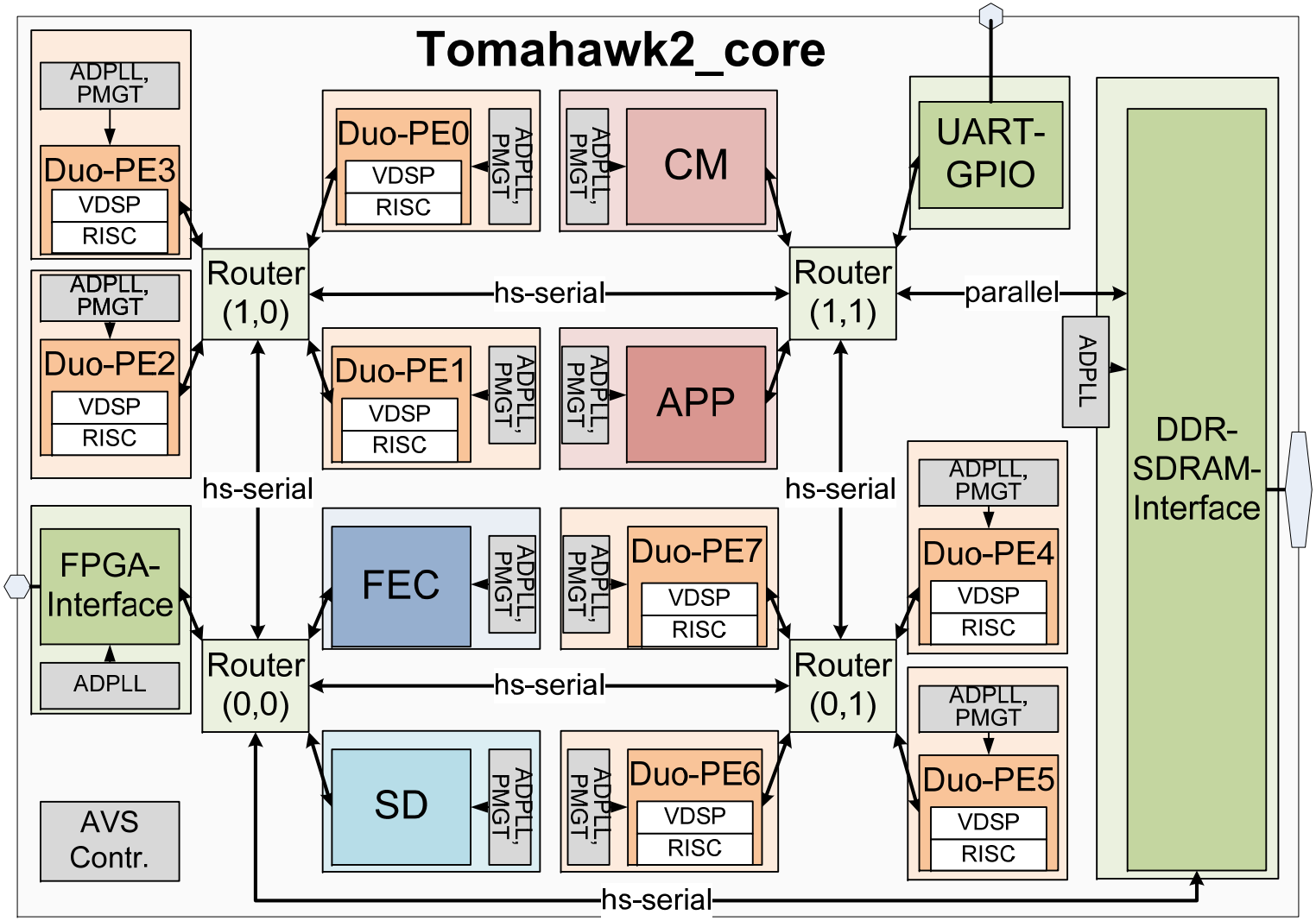
Challenges:

- ➔ Variable application workloads and start times
- ➔ Non-deterministic task execution times
- ➔ Latency and throughput requirements
- ➔ Heterogeneous architecture (e.g., for signal processing)
- ➔ Power and data management
- ➔ Energy-aware scheduling

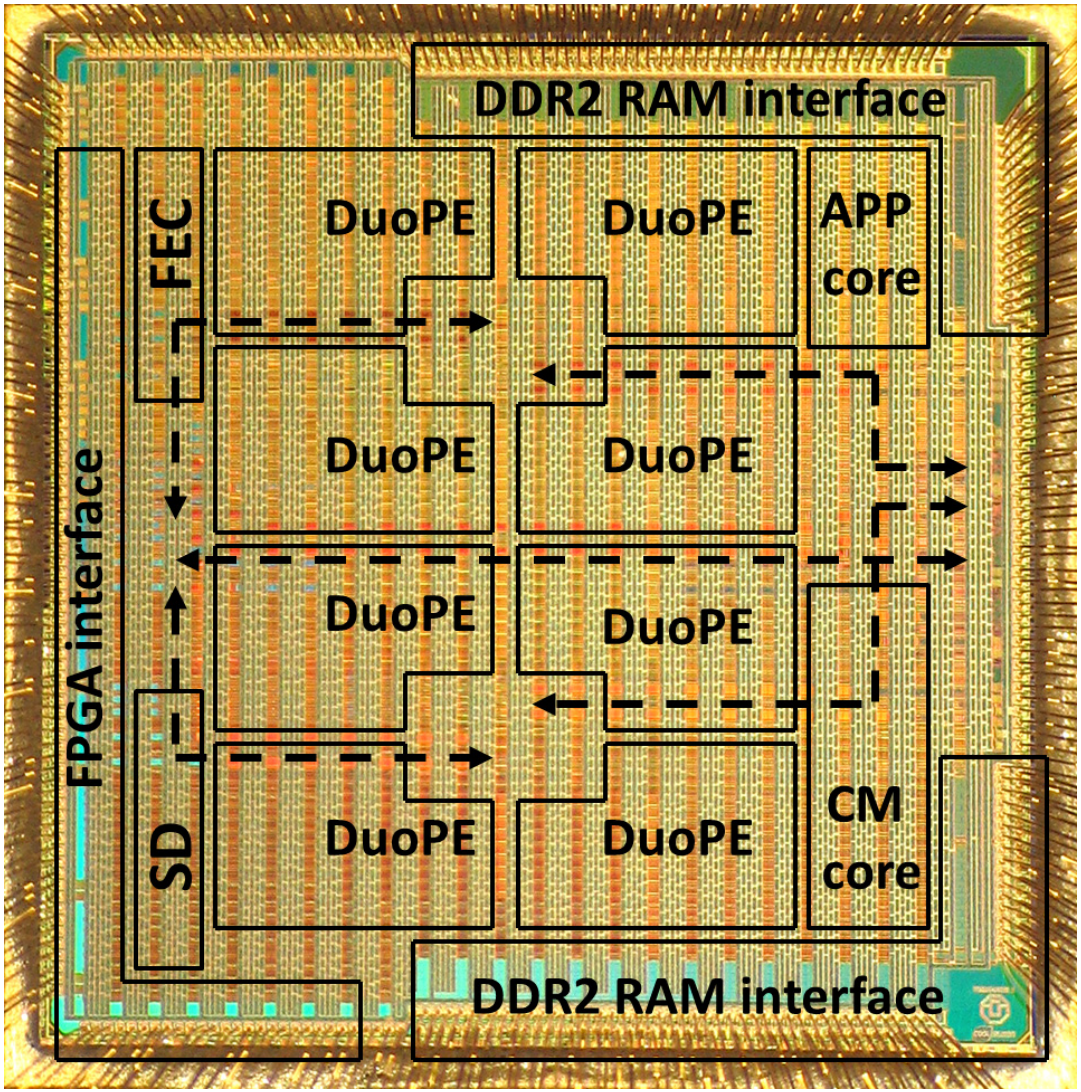
Tomahawk Architecture Framework



Block Diagram



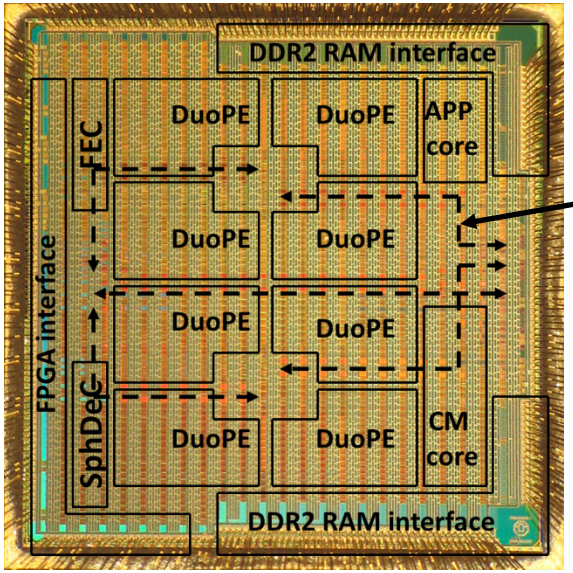
Tomahawk 2 Die Photo

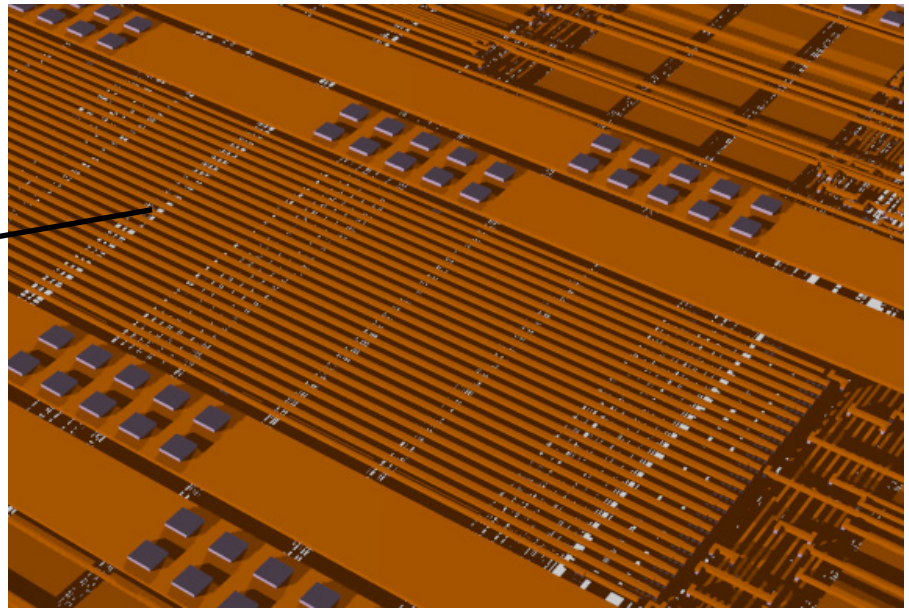


– – serial NoC link

High-speed Serial NoC Links

- Point-to-point connections within star-mesh NoC
- 80 Gbit/s (8GBit/s/lane) serial links [1] at <150fJ/Bit/mm
- Compact floorplan realization by bridging of core macros

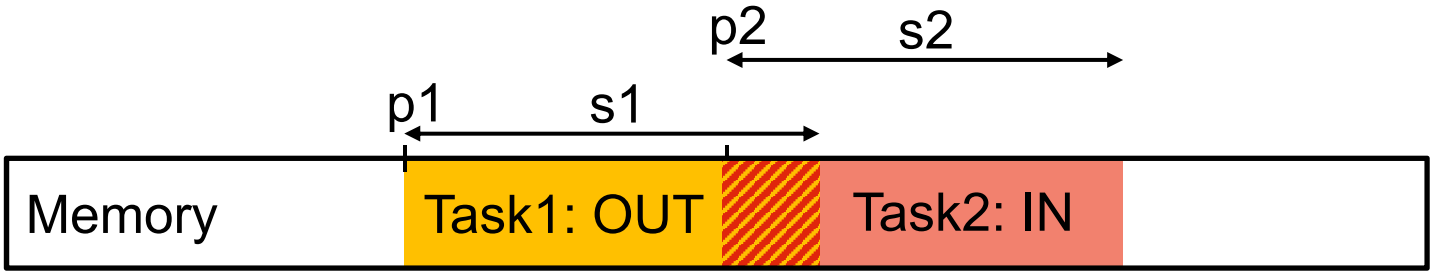




[1] D. Walter, S. Höppner, H. Eisenreich, G. Ellguth, S. Henker, S. Hänzsche, R. Schüffny, M. Winter, G. Fettweis, A Source-Synchronous 90Gb/s Capacitively Driven Serial On-Chip Link Over 6mm in 65nm CMOS, ISSCC Dig. Tech. Papers, 2012

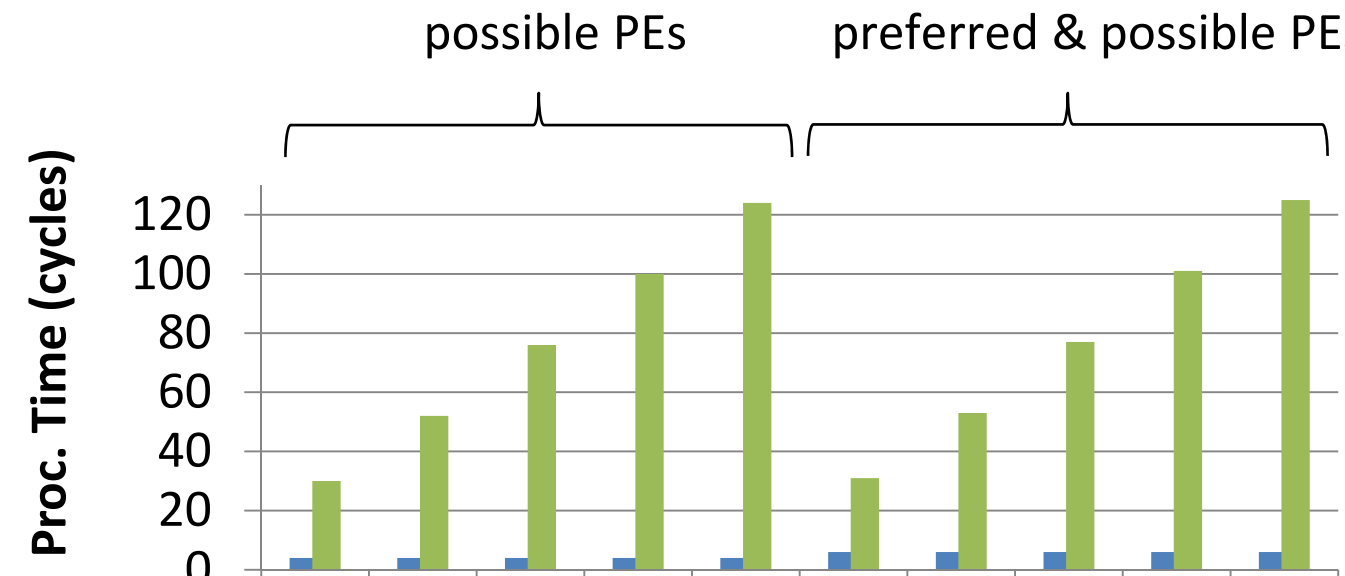
Platform Control

- **CoreManager ➔ ASIP**
  - Central task scheduling unit
  - Data plane control (8xVDSP, 8xRISC, 1xFEC, 1xSD)
  - Scheduling-specific instruction set ➔ fully flexible in contrast to ASIC
  - Optional: dynamic data dependency checking



CoreManager: Dynamic PE Allocation

- **Heterogeneous MPSoC ➔ several types of cores**



PE-ID	0	7	15	23	31	0	7	15	23	31
CM-ASIP	4	4	4	4	4	6	6	6	6	6
CM-RISC	30	52	76	100	124	31	53	77	101	125

[2] O. Arnold, B. Noethen, G. Fettweis, Instruction Set Architecture Extensions for a Dynamic Task Scheduling Unit, *IEEE Computer Society Annual Symposium on VLSI 2012 (ISVLSI)*, pp.249,254, 19-21 Aug. 2012

CoreManager Comparison

	ASIC [3]	RISC	ASIP (this work)
Scheduling Configurability	Fixed	Flexible	Flexible
Task Queue size	16	16-256	16-256
Max. Frequency [MHz]	175	445	445
Task scheduling [us]	0.4	16.9	0.9
Technology [nm]	130	65	65
Supply Voltage [V]	1.3	1.2	1.2
Power [mW] @fmax	282	68	74.6
Energy per Task [nJ] @fmax	113 (27*)	1149	67
Area (logic) [mm²]	4.51 (1.13*)	0.34	0.49
ATE product [mm²*us*nJ]	204 (12*)	6602	29

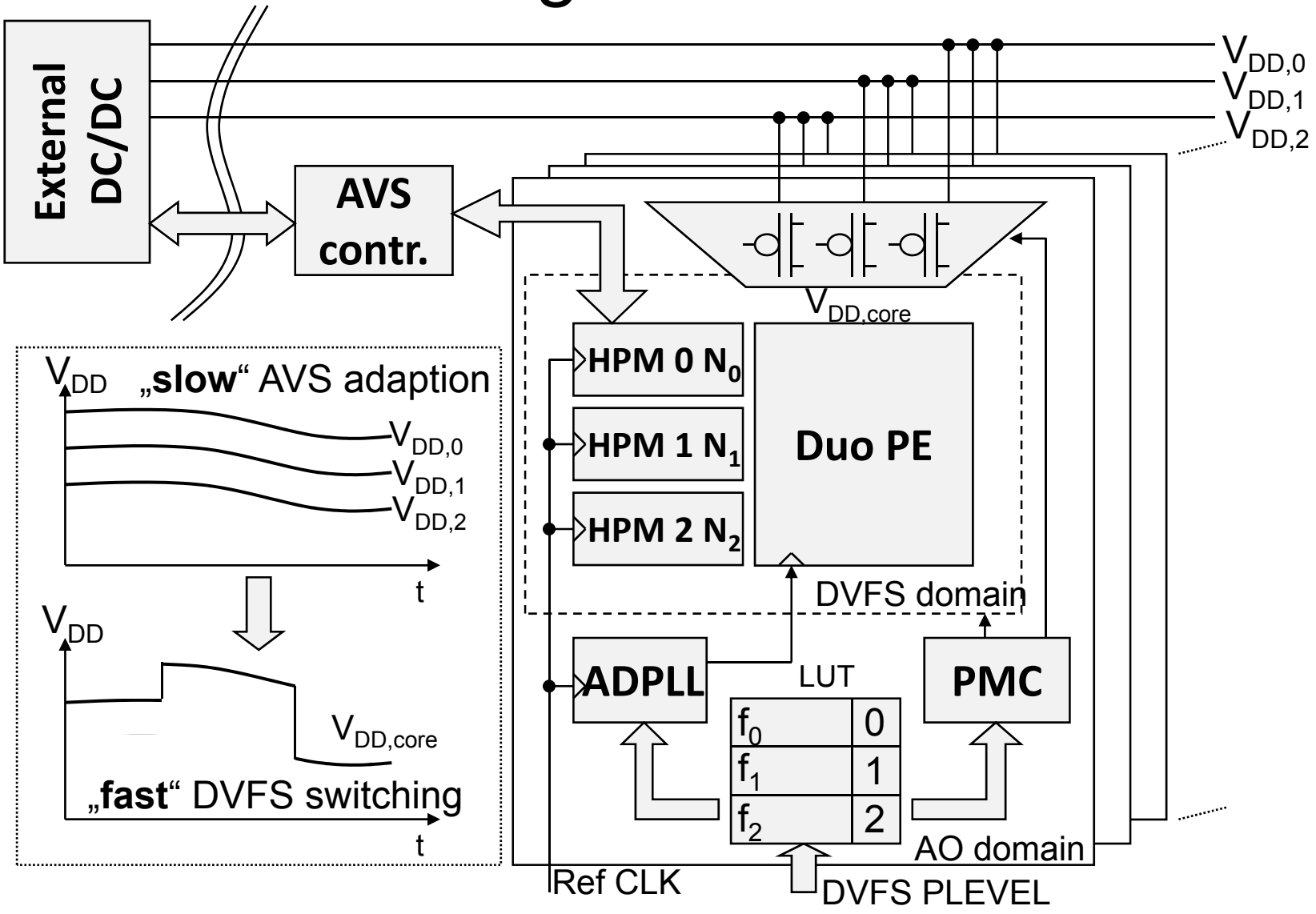
*scaled to 65nm @1.2V

[3] T. Limberg et al., A Fully Programmable 40 GOPS SDR Single Chip Baseband for LTE/WiMAX Terminals, in Proceedings of the 34th European Solid-State Circuit Conference (ESSCIRC'08), Edinburgh, UK, 15.9. - 19.9.2008

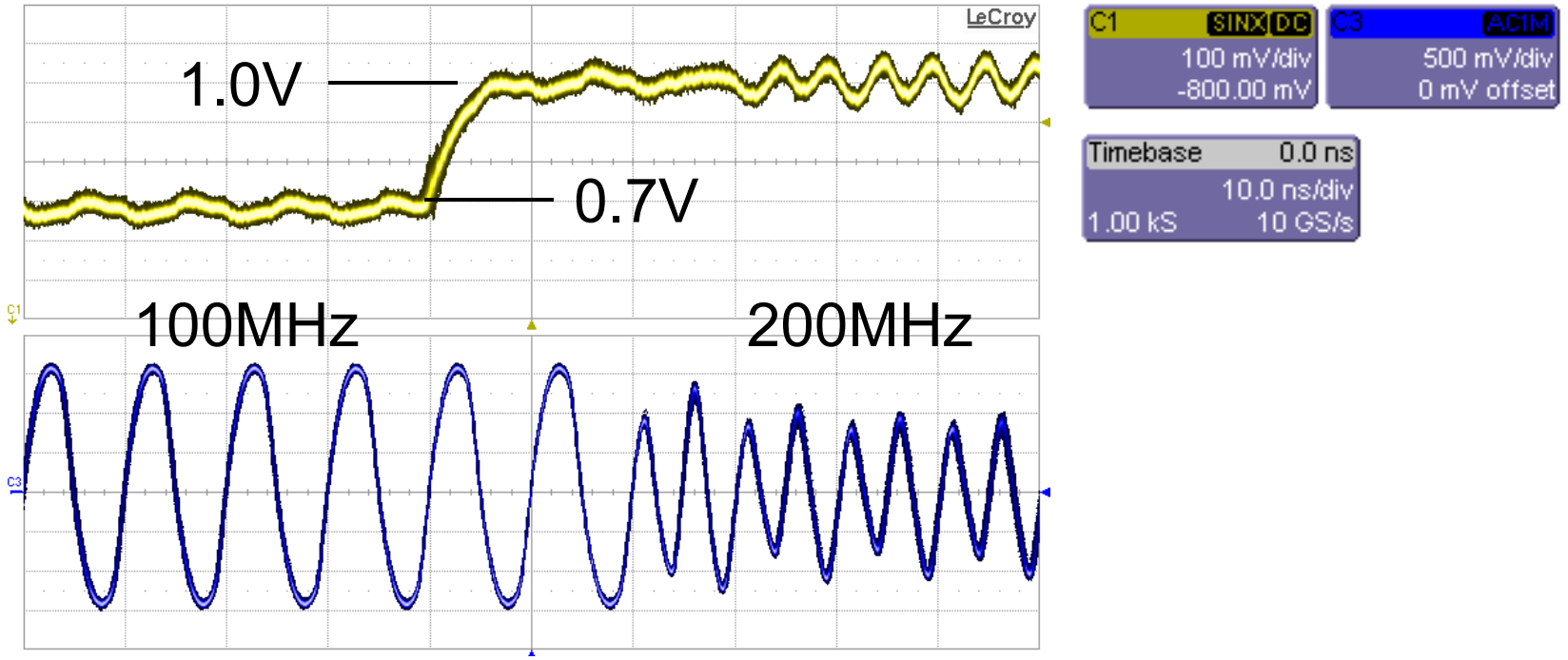
Power Management

- **Fine-grained: CoreManager**
  - Based on application requirements (e.g., start times and deadlines) and system status (e.g., allocated PEs)
  - Result: target frequency and voltage level for each PE for each task
  - Enabled by fine grained **fast** DVFS on PE level
- **Coarse-grained**
  - Global AVS ➔ voltage control of DVFS levels
  - Result: **slow** voltage adaption for temperature and process for DVFS rails

Power Management Architecture



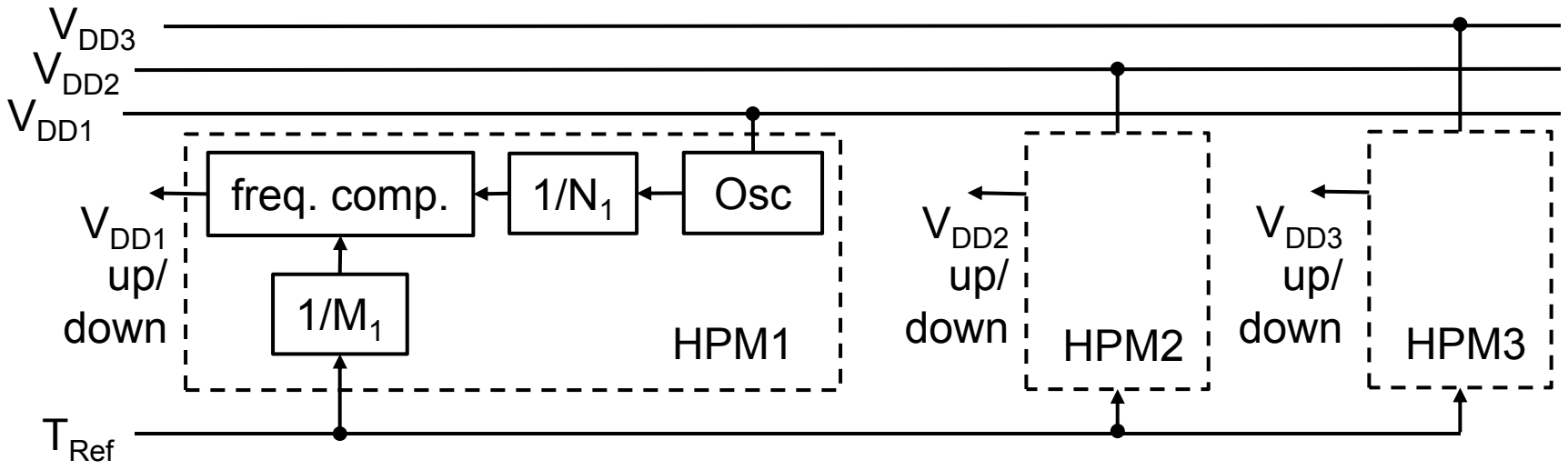
Fast DVFS Measurement Result



- Ultra fast DVFS level change of Duo-PE in <20ns

3-Level DVFS with combined AVS

- 3 performance levels (V_{DD1},f_1) (V_{DD2},f_2), (V_{DD3},f_3)
  - **Fast** DVFS switching between rails
- Hardware Performance Monitor (HPM) with virtual frequency multiplication capability
  - Configurable ring oscillator replicates critical timing
  - **Slow** AVS adaption of V_{DD} such that $f_{Osc,i}=f_{Ref}\cdot N_i/M_i=f_i$

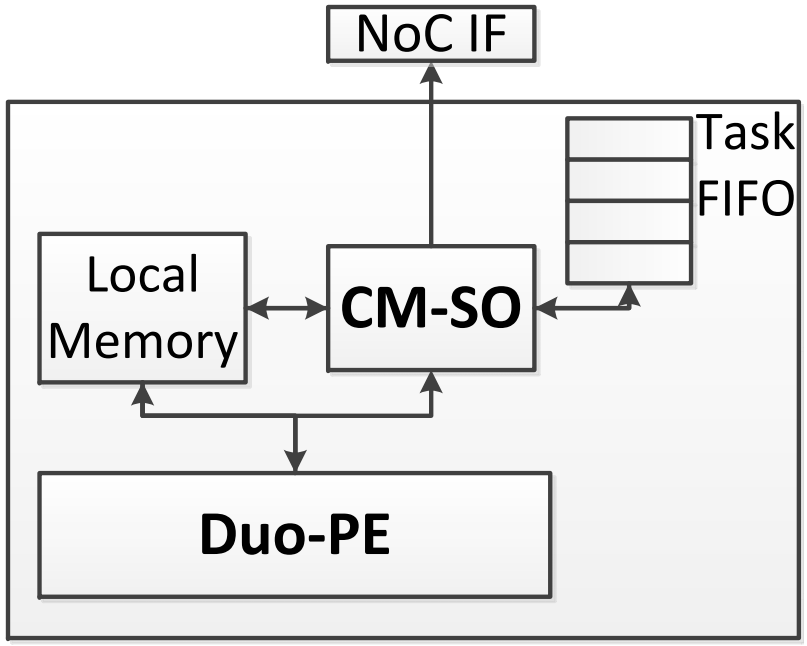


Duo-PE Concept

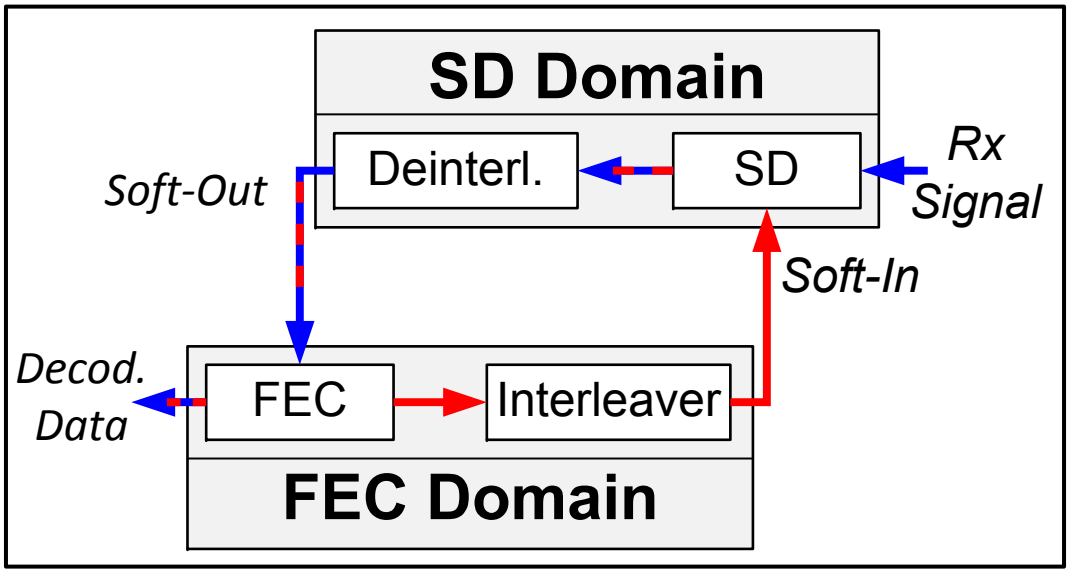
- Motivation:
  - Data transfer among heterogeneous cores increases power consumption, network congestions and latencies
- Heterogeneous cores connected to local Memory:
  - 4-fold SIMD Vector DSP → 16-bit fixed point comp.
  - General Purpose RISC Core → high precision FP comp.
- Advantage:
  - Further increase of data locality and area efficiency
  - Dynamic core selection during runtime
  - Use of existing compilers

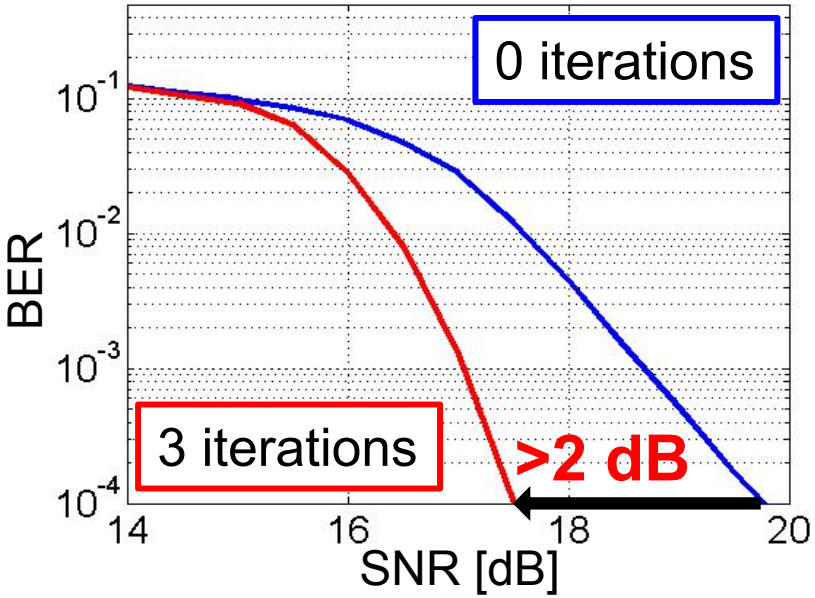
Duo-PE

- **Duo-PE controlled by the CM-Spinoff**
  - DMA controller programmed by task descriptions
  - Allows concurrent data prefetching during task execution
  - Activates the clock for the requested core



Iterative MIMO Detection-Decoding

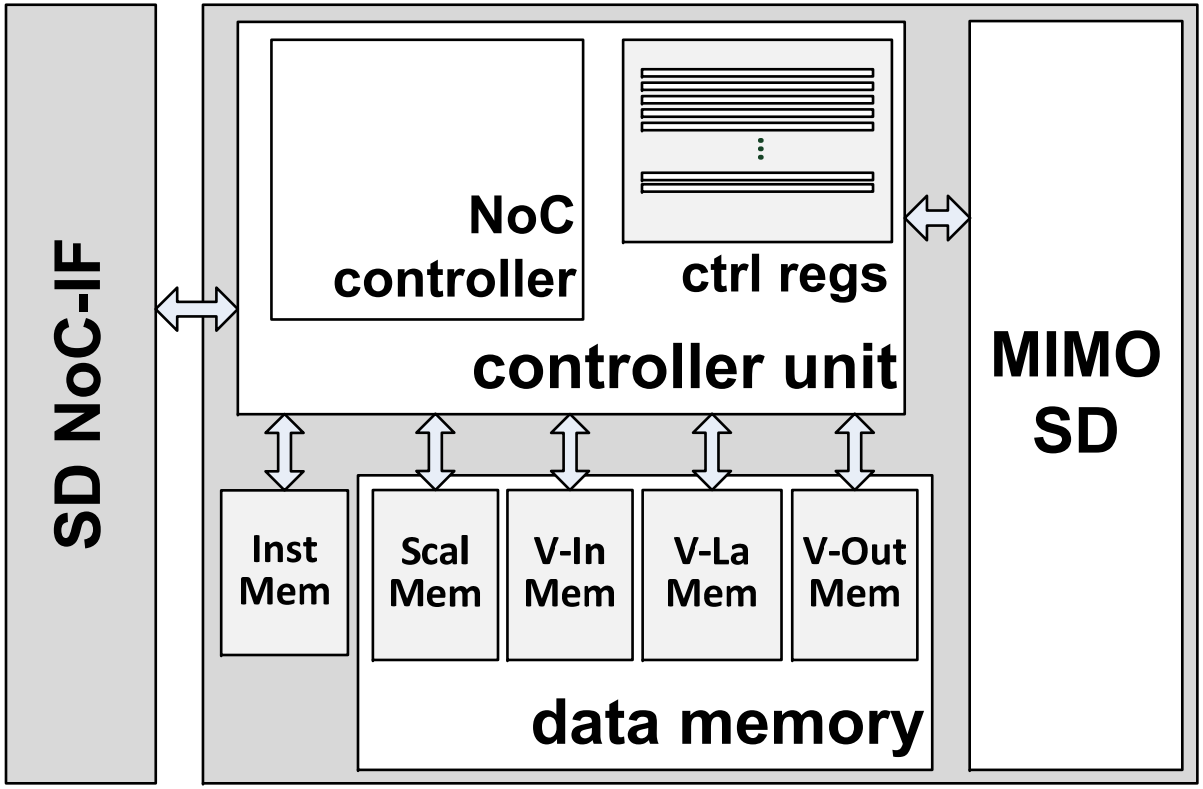




- Exchange of soft-bit information between **Sphere Detector** (SD) and **Forward Error Correction** (FEC)
- Improvement of communications performance compared to non-iterative *Tommy*[4] chip due to turbo iterations [5]

[4] M. Winter, et al., "A 335Mb/s 3.9mm2 65nm CMOS flexible MIMO detection-decoding engine achieving 4G wireless data rates," ISSCC Dig. Tech. Papers, pp.216-218, 2012.
[5] E. P. Adeva, et al., "VLSI Architecture for MIMO Soft-Input Soft-Output Sphere Detection", J. of Signal Processing Systems, vol. 70, is. 2, pp. 125-143, 2013.

Soft-In Soft-Out Sphere Detector

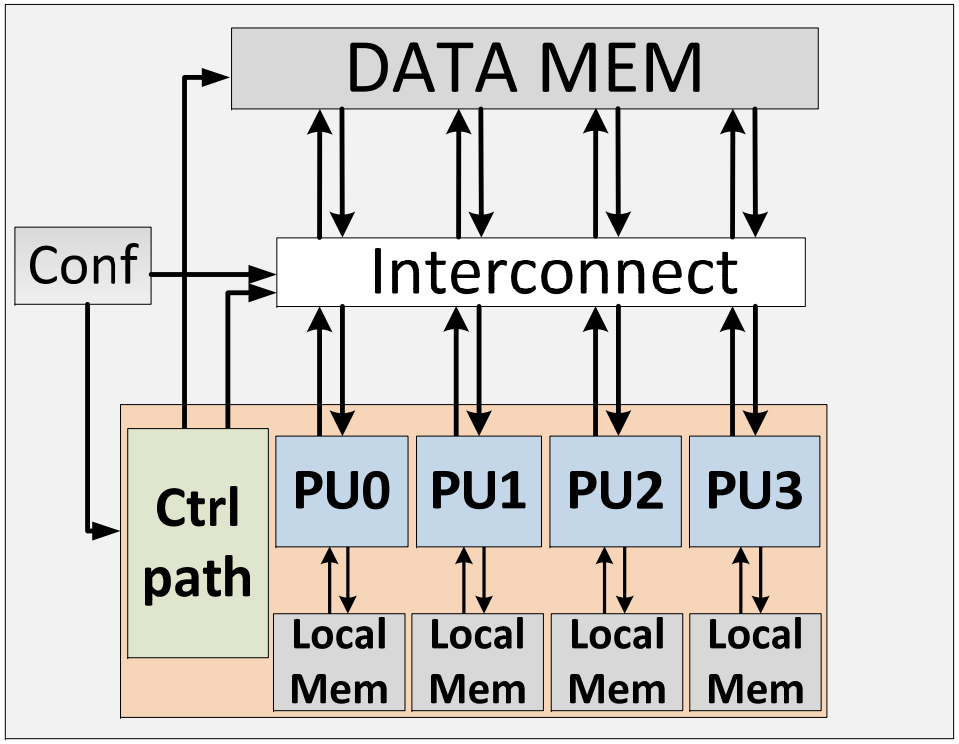


f_{max}	445 MHz
Area	0.52 mm²
Power*	87 mW
Through-put	396 Mb/s

*@1.2V

- ASIP design, controlled by VLIW
- Tree-search based MIMO detection algorithm
- Extended *Tommy* SD ⇨ Processing of Soft-In (FEC output)

Multi-Mode Forward Error Correction



f_{max}	500 MHz
Area	1.15 mm²
Power*	360 mW
Throughput**	155 Mb/s

* @1.2V
** LDPC

- Programmable ASIP implementation
- Four SIMD-parallel FEC processing units (PUs)
- Combined decoding operations for Viterbi/Turbo/LDPC codes

Tomahawk2 Components

		Area [mm2]		f_{max} [MHz] @VDD=1.2 V	Throughput @f_{max}	$P_{Application\text{-}Scenario}$* [mW]
		total	*mem*			
APP		0.582	0.245	445	890 MOPS	off
CM		1.360	0.870	445	1.1 MTasks/s	14.1 @200MHz, 0.9 V
Duo-PE	**RISC**	1.357	0.800	445	890 MOPS	off
	VDSP			500	10 GOPS	35.0 @282 MHz, 0.9 V
SD		0.522	0.260	445	396 Mb/s	36.5 @200 MHz, 1.15 V
FEC		1.154	0.618	500	155 Mb/s	132.2 @200 MHz, 1.15 V
FPGA-IF		0.602	-	500	10 Gb/s	-
DDR-IF		4.552	-	400	12.8 Gb/s	-
NoC		3.417	-	500	80 Gb/s/link	18.0 @286 MHz, 1.15 V

*4×4 MIMO 3GPP-LTE baseband application

Demonstrator Setup

SDR Architectures for MIMO 3GPP-LTE/WiMAX

	Tomahawk2	Magali [6]	Tomahawk [3]
Clocking and Power Management	GALS, local DVFS & AVS, power-gating	GALS, local DFS	Global clock
Scheduling	Dynamic (flexible, energy adaptable)	Static	Dynamic (fixed algorithm)
Peak Performance	105 GOPS (3.6 GFLOPS)	37 GOPS	40 GOPS
Application for power measurements	4x4 MIMO 3GPP-LTE Rx baseband, 60 Mbit/s	4x2 MIMO, 2x2 MIMO Tx, MAC, 10.8 Mbit/s	LTE/WiMAX
Power consumption	480 mW @ 1.15V	477 mW @ 1.2V	1.2W @ 1.3V
NoC Throughput (per link)	80 Gbit/s	17 Gbit/s	5.47 Gbit/s
Die size	36 mm²	29.6 mm²	100 mm²
Technology	65 nm	65 nm	130 nm

[6] Clermidy, *et al*., A 477mW NoC-based digital baseband for MIMO 4G SDR, *ISSCC Dig. Tech. Papers,* pp. 278-279, 2010

Conclusion

- Heterogeneous MPSoC with dynamic task-scheduling
  - Several types of PEs: VDSPs, GP-cores and ASIPs
  - Data-plane control: CoreManager with scheduling-specific instruction set
  - Star-Mesh NoC with serial links
  - Dynamic and Adaptive Power Management
  - Iterative Detection-Decoding for increased communication performance

A Multi-Standard 2G/3G/4G Cellular Modem Supporting Carrier Aggregation in 28nm CMOS

Michael Breschel[1], Peter Almers[1], Fredrik Angsmark[1], Alberth Arvidsson[1], Harald Bauer[2], Kees van Berkel[3], Joaquin Canovas[1], Minh Do[1], Anders Ekelund[1], Torsten Larsson[1], Bo Lincoln[1], Magnus Malmberg[1], Masao Naruse[4], Masashi Onishi[4], Christer Östberg[1], Jean-Paul Smeets[3], Mario Vergara Escobar[1], Juergen Voelkl[2], Emma Wittenmark[1]

[1]Ericsson, Lund, Sweden, [2]Ericsson, Nuremberg, Germany, [3]Ericsson, Eindhoven, The Netherlands, [4]Ericsson, Yokohama, Japan

DB7450R = multi-standard digital baseband IC

The supported *Radio Access Technologies* (RAT) include:

RAT	DL category	Max DL bitrate	UL category	Max UL bitrate
LTE TDD/FDD	4	150 Mb/s	4	50 Mb/s
LTE CA up to 30 MHz	4	222 Mb/s	4	50 Mb/s
WCDMA	28	84 Mb/s	8	11 Mb/s
TD-SCDMA	15	2.8 Mb/s	6	2.2 Mb/s
GSM EDGE	33	296 kb/s	33	236 kb/s

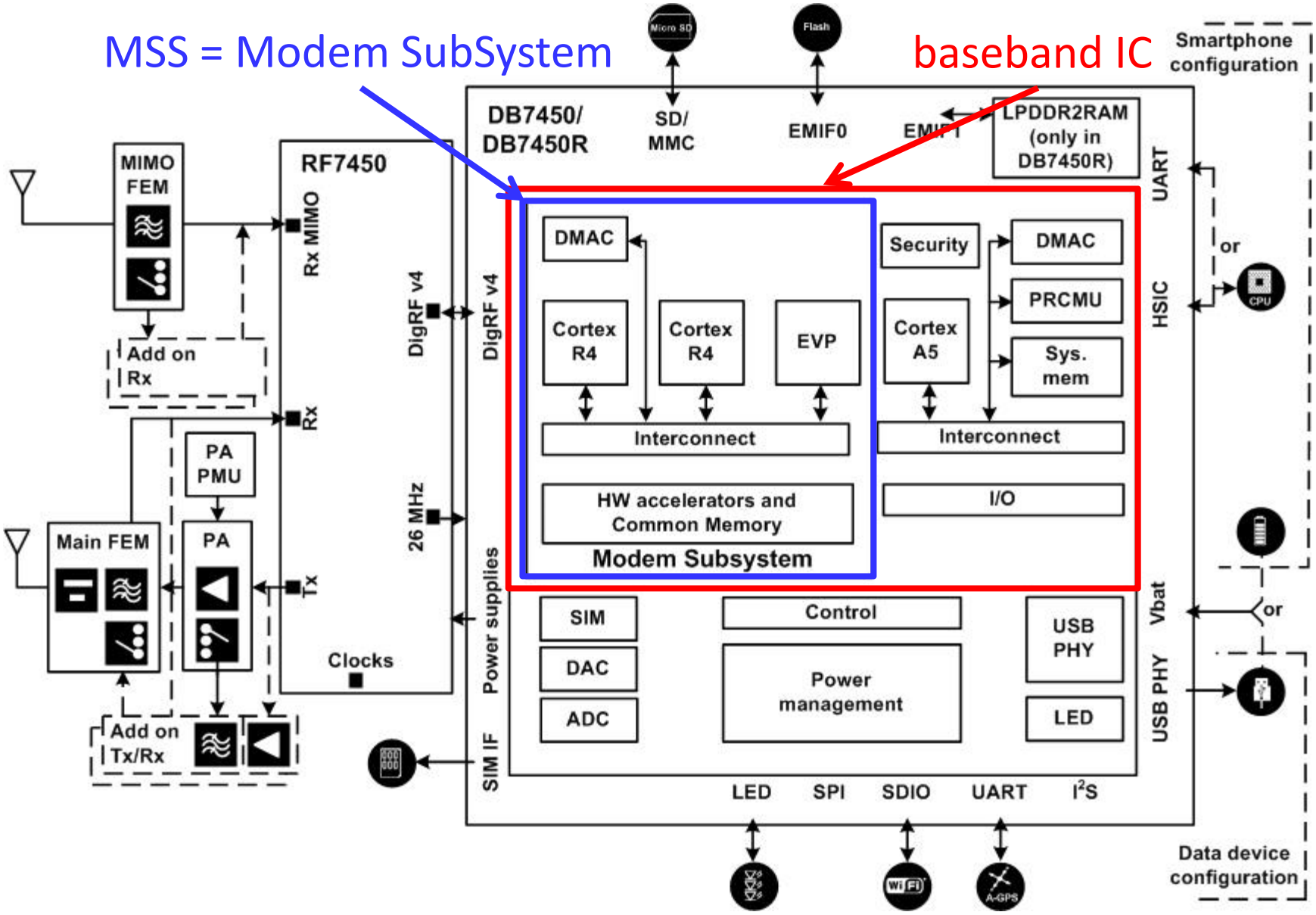
LTE carrier aggregation

- Combination of carriers at different frequencies to obtain a larger effective bandwidth
  - For example 10+10 MHz, 5+15 MHz or 10+20 MHz
- Relevant for network operators with limited contiguous spectrum allocation
- The carriers are treated independently at the physical layer and combined to one data stream at a higher layer in the protocol

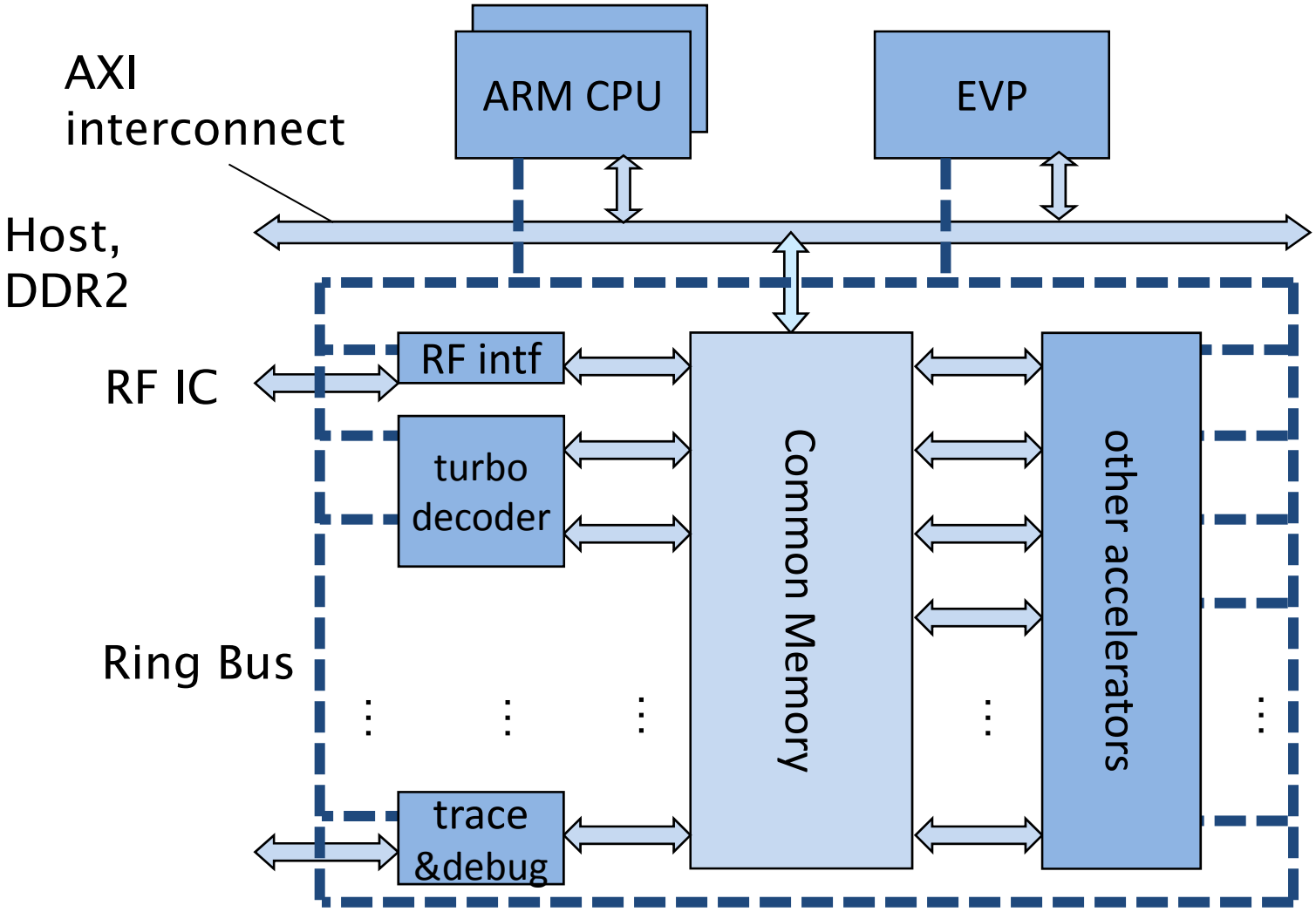
DB7450R system-in-package

- Digital baseband die implemented in 28 nm CMOS technology
  - The modem sub system (MSS) inside the baseband supports multiple RATs with a unified HW and SW architecture
- Power management unit die
- LPDDR2 SDRAM die

DB7450R baseband IC in context



MSS hardware architecture



Hardware reuse principle

- All programmable cores are shared
- Memory is handled in a pool and is reused between RATs
- Accelerators are generic and non-RAT specific where possible
- Common control principle using the ring bus
- The reuse reduces the total die area for a multi RAT capable modem

Modem Subsystem: Memory

On-chip Common Memory (CM) is used for data exchange between accelerators

- Allows flexible allocation of memory per RAT
- 2x16 banks at 260 MHz interleaved at 16bit word level offers a predictable bandwidth and latency
- Total bandwidth 133 Gbit/s
- The latency can be afforded because the CM is primarily used for large batch oriented accesses
- CPU and embedded vector processor (EVP) access CM via an AXI bridge

Ring bus concept

- All processors and accelerators connected to the CM are connected by a ring bus (RB) – a common multi-master interconnect to exchange messages between the hardware units
- All control information is carried on the RB
- The RB enables chaining of multiple hardware processing tasks as well as software processing tasks (e.g. on the EVP) without any control software steps in between

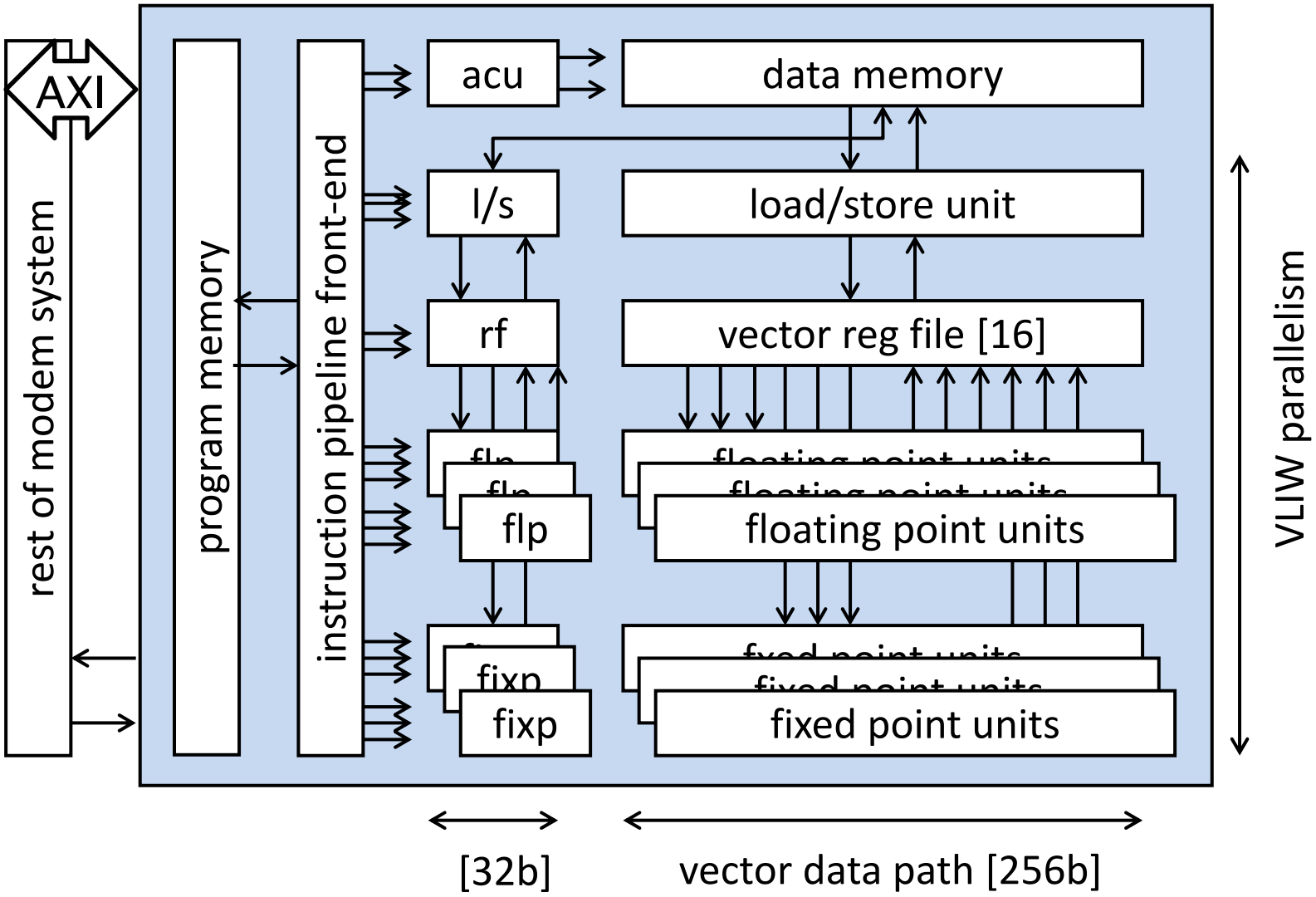
Ring bus implementation

- Each accelerator includes a configurable control interface that contains the RB interface and mail boxes for the messages
- Messages have a fixed 128 bit length with a header specifying type of message and destination
- The payload typically specifies a task to be executed and a pointer to CM
- The control interface governs the execution of a task according to a list of messages stored in CM
  - Besides setting control parameters and initiating the HW function, the messages can be forwarded to other clients on the RB

Multi RAT hardware flexibility

- The uniform and scalable architecture allows a flexible hardware software partitioning
  - This gives the system designer the flexibility to find the optimum hardware software tradeoff (area, power, late product changes, …)
- Illustration of the flexibility of the architecture:
  - Support of 10+20 MHz carrier aggregation
  - Requires the re-scheduling of the signal processing on the accelerators and processors
  - No hardware changes are needed

EVP architecture

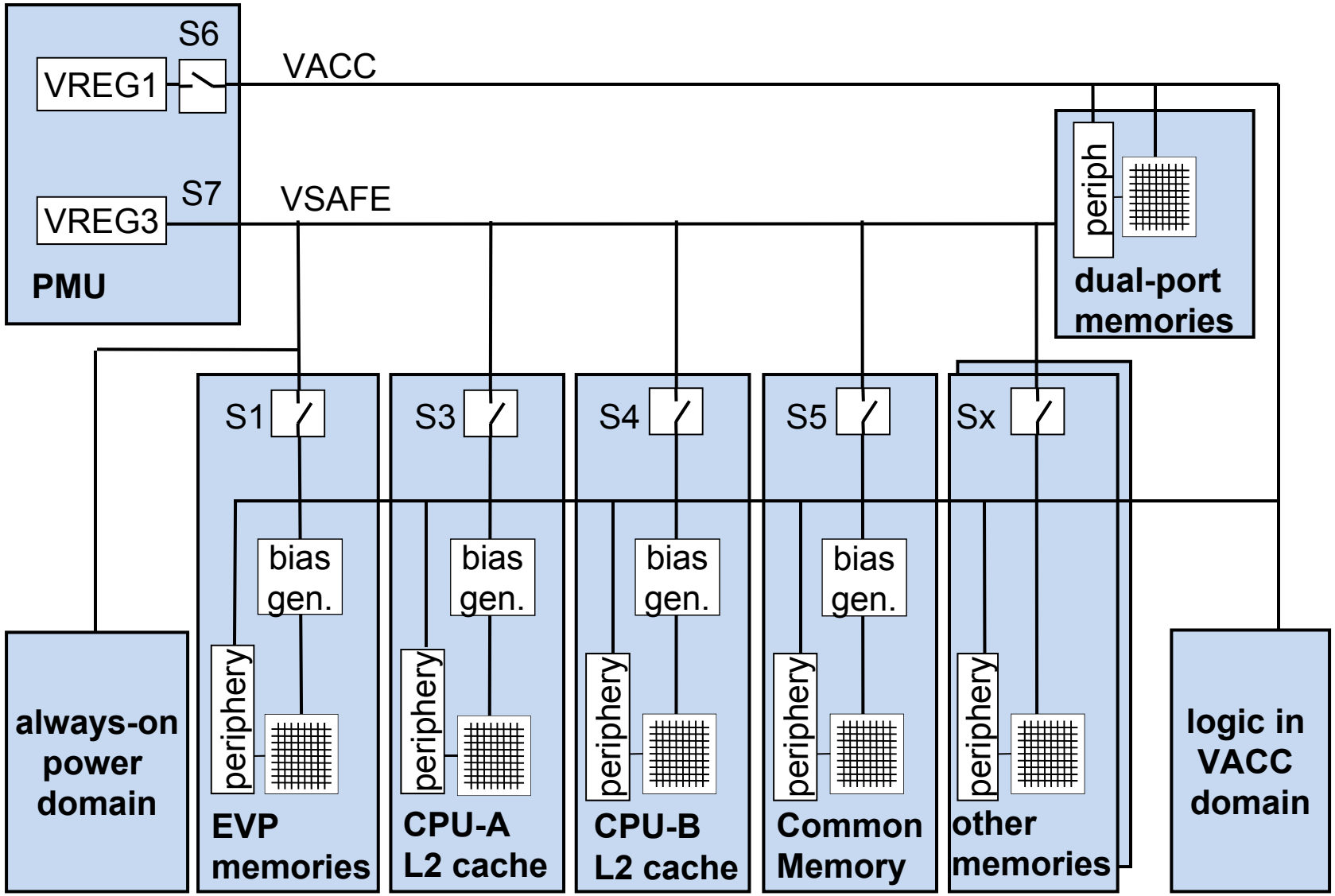


Embedded Vector Processor (EVP)

Programmable DSP supporting up to 100 operations per cycle at 416 MHz

- SIMD architecture with 256 bit vectors
- Integrated scalar processor operating in parallel
- Using VLIW from a unified program stream
- Floating point support with 32+4 parallel multipliers
- Supporting 8+1 complex floating point MACs per cycle
- Low power consumption (below 0.5 mW/MHz) including data and program memories

DB7450/MSS power architecture



Power domains

- Always-on domain less than 1% of die area
  - The only powered domain during sleep/flight mode
  - Uses 1 MHz and 32768 Hz clock during sleep mode
- Separate power domain for the RF interface
  - Powered only during active use
- The rest of the die is powered in active mode
  - HW controlled switch between active and sleep mode

DVFS and AVS

- Dynamic Voltage and Frequency Scaling is applied in active mode
  - Dependent on current performance requirements
  - Switching done at pre-defined trigger levels
  - Dynamic switching aligned to the RAT timing
    - Periodic performance increase to meet dead lines while average load is relatively low
- Adaptive Voltage Scaling (AVS) is used on each device to adapt to its process characteristics

Clock gating

- Extensive clock gating in 4 levels
  - At PLL outputs
  - At central clock control unit
  - At the block interface
  - To specific functions within a block
- Clock gating is hardware controlled and handled locally by the blocks
- The interconnect network is segmented and transaction based clocking is enabling the clock only to the parts that need it

Power consumption

- Measured at approximately 25°C
- Supply voltage 1.03V or 0.90V depending on use case
- Ongoing software improvements and optimizations are not included
- Lowest published numbers for a 4G modem

DBB power consumption [mW]

for selected use cases

RAT	downlink category	downlink bandwidth [MHz]	downlink max bitrate [Mb/s]	uplink max bitrate [Mb/s]	power consumption [mW]
GSM idle DRX9					1.0
HSPA	14		21		183
LTE	3	10	63	25	321
LTE	3	20	100	50	400
LTE	4	20	150	50	430
LTE	4	10+10	150	25	438

Conclusion

- Unified HW/SW architecture for cellular access:
  - enables resource sharing among RATs
  - supports scaling in number of carriers, bandwidth, etc. (M7450: carrier aggregation up to 10+20 MHz)
  - is programmable where it counts
  - yields lowest published power numbers for a 4G modem
- The architecture is prepared for future evolution of the 3GPP standard, including:
  - heterogeneous networks
  - more advanced receiver algorithms for improved link performance